Eastern
Economy
Edition

Electronic Instruments and Instrumentation Technology

M.M.S. Anand

PHI

Electronic Instruments and Instrumentation Technology

M.M.S. ANAND
Professor
Electronics and Instrumentation
and Registrar
Birla Institute of Technology & Science
Pilani

PHI Learning Private Limited
Delhi-110092
2005

ELECTRONIC INSTRUMENTS AND INSTRUMENTATION TECHNOLOGY
M.M.S. Anand

© 2004 by PHI Learning Private Limited, Delhi. All rights reserved. No part of this book may be reproduced in any form, by mimeograph or any other means, without permission in writing from the publisher.

ISBN-978-81-203-2454-1

The export rights of this book are vested solely with the publisher.

Ninth Printing ... ... January, 2013

Published by Asoke K. Ghosh, PHI Learning Private Limited, Rimjhim House, 111, Patparganj Industrial Estate, Delhi-110092 and Printed by Mohan Makhijani at Rekha Printers Private Limited, New Delhi-110020.

I dedicate this book to my parents.

Contents

Preface .... *xv*

Section I—Instruments and Accessories

1. ANALOG METERS **3–53**
1.1 Introduction 3
1.2 Permanent Magnet Moving Coil (PMMC) Meters 3
1.3 DC Analog Meters 10
    1.3.1 DC Ammeter 10
    1.3.2 DC Voltmeter 16
    1.3.3 Ohmmeter 20
    1.3.4 Multimeter 25
    1.3.5 Calibration 27
1.4 AC Analog Meters 28
    1.4.1 Half-wave Rectifier-based Meter 28
    1.4.2 Full-wave Rectifier-based Meter 32
    1.4.3 Nonsinusoidal Waveform Input 33
    1.4.4 Peak Reading Voltmeter 34
    1.4.5 Peak-to-Peak Reading Voltmeter 37
1.5 Other Meter Movements 38
    1.5.1 Electrodynamic Meter Movement 38
    1.5.2 Moving Iron (Vane) Meter Movement 41
1.6 Electronic Analog Meters 43
    1.6.1 Electronic Analog DC Voltmeter 43
    1.6.2 Electronic Analog DC Ammeter 45
    1.6.3 Electronic Analog AC Voltmeter 45
    1.6.4 Electronic Analog AC Ammeter 47
    1.6.5 Electronic Analog Ohmmeter 47
    1.6.6 Electronic Analog Multimeter 48
Questions *49*

2. DIGITAL METERS 54–118
2.1 Introduction 54
2.2 Digital-to-Analog Converter 55
  2.2.1 Transfer Characteristics 56
  2.2.2 D/A Conversion Techniques 57
  2.2.3 Digital Input Modes of Operation 61
  2.2.4 Performance Characteristics of D/A Converters 61
2.3 Analog-to-Digital Converter 66
  2.3.1 Transfer Characteristics 66
  2.3.2 A/D Conversion Techniques 69
2.4 Digital Meter Displays 92
2.5 Specifications of Digital Meters 92
  2.5.1 Display Digits and Counts 92
  2.5.2 Resolution 93
  2.5.3 Sensitivity 94
  2.5.4 Accuracy, Uncertainty and Repeatability 94
  2.5.5 Speed and Settling Time 95
  2.5.6 Average/True Root Mean Square (TRMS) 95
  2.5.7 Crest Factor 97
  2.5.8 Form Factor 97
  2.5.9 Normal Mode Rejection Ratio (NMRR) 97
  2.5.10 Common Mode Rejection Ratio (CMRR) 98
  2.5.11 Effective Common Mode Rejection Ratio (ECMRR) 98
  2.5.12 Zero or Offset 98
  2.5.13 Frequency Response 99
  2.5.14 Input Resistance 99
  2.5.15 Input Bias Current 99
  2.5.16 Overload Protection 99
  2.5.17 Autoranging 100
  2.5.18 Autozero 100
  2.5.19 Four-wire Remote Sensing 100
  2.5.20 Environmental Specifications 101
  2.5.21 Safety Ratings 102
  2.5.22 Electromagnetic Compatibility (EMC) 103
  2.5.23 Additional Measurement Features 103
2.6 DMM Calibration 105
  2.6.1 DC Voltage Calibration Strategy 105
  2.6.2 Functional Tests 106
  2.6.3 Input Characteristics 106
  2.6.4 Resistance Calibration Strategy 107
  2.6.5 Resistance Function Linearity Tests 107
  2.6.6 AC Voltage Calibration Strategy 108
  2.6.7 Current Calibration Strategy 109
  2.6.8 Zero Considerations 110
  2.6.9 Generating a Test Plan 110
  2.6.10 Measurement Uncertainty 111
2.7 Selecting a DMM 112
Questions 118

3. ANALOG AND DIGITAL OSCILLOSCOPES 119–158
3.1 Introduction 119
3.2 Types of Oscilloscopes 120
3.3 Analog Oscilloscopes 122
3.4 Digital Storage Oscilloscope (DSO) 132
3.5 Digital Phosphor Oscilloscope (DPO) 137
3.6 The Controls 138
    3.6.1 Display Controls 138
    3.6.2 Vertical Controls 139
    3.6.3 Horizontal Controls 143
    3.6.4 Trigger Controls 144
    3.6.5 Acquisition Controls for Digitizing Oscilloscopes 150
    3.6.6 Sampling Methods 151
    3.6.7 Other Controls 151
3.7 Measurements 151
    3.7.1 Voltage Measurements 152
    3.7.2 Time and Frequency Measurements 153
    3.7.3 Pulse and Rise Time Measurements 153
    3.7.4 Fully Automated Measurements with Digitizing Oscilloscopes 154
3.8 Selecting an Oscilloscope 155
    3.8.1 The “5 Times” Rule for Bandwidth 155
    3.8.2 Sample Rate 155
    3.8.3 Triggering 156
    3.8.4 Record Length 156
3.9 Analog vs DSO vs DPO 156
Questions 157

4. PROBES 159–178
4.1 Introduction 159
4.2 Types of Probes 160
    4.2.1 Passive Voltage Probes 160
    4.2.2 Active Voltage Probes 163
    4.2.3 Current Probes 164
    4.2.4 Logic Probes 166
    4.2.5 Optical Probes 166
    4.2.6 Other Probes 166
4.3 Probe Loading and Measurement Effects 167
    4.3.1 Resistive Loading 167
    4.3.2 Capacitive Loading 168
    4.3.3 Inductive Loading 172
4.4 Probe Specifications 173
Questions 177

5. BRIDGE INSTRUMENTS 179–225
5.1 Introduction 179
5.2 Direct Current Bridges 179
  5.2.1 Wheatstone Bridge 180
  5.2.2 Kelvin Bridge 195
5.3 Alternating Current Bridges 200
  5.3.1 Maxwell Bridge 205
  5.3.2 Hay's Bridge 206
  5.3.3 Schering Bridge 208
  5.3.4 Wien Bridge 210
  5.3.5 Wagner Ground Connection 211
5.4 Anderson Loop 212
5.5 LCR Bridge 218
Questions 221

6. RECORDERS AND DATA LOGGERS 226–240
6.1 Introduction 226
  6.1.1 Distinction between a Recorder and a Data Logger 227
6.2 Strip Chart Recorder 228
  6.2.1 PMMC Galvanometer Type 228
  6.2.2 Potentiometric/Servo/Null Balancing Recorders 232
6.3 X-Y Recorder 236
6.4 Data Loggers 237
6.5 Videographic or Paperless Recorders 238
6.6 Recorder Selection 239
Questions 240

7. SIGNAL GENERATORS 241–294
7.1 Introduction 241
7.2 Sine-wave Generators 242
  7.2.1 Audio Frequency Signal Generators 244
  7.2.2 Radio Frequency Signal Generators 247
7.3 Non-sinusoidal Generators 254
7.4 Function Generators 264
7.5 Pulse Parameter Definitions 265
7.6 Frequency Synthesis Techniques 269
  7.6.1 Direct Analog Synthesis 270
  7.6.2 Indirect Synthesis 272
  7.6.3 Direct Digital Synthesis 277
7.7 Digital Signal Generators 282
  7.7.1 Arbitrary Waveform Generator 283
  7.7.2 Arbitrary Function Generator 288
  7.7.3 Data Generator 289
7.8 Key Characteristics of Digital Signal Generators 290
Questions 292

8. SIGNAL ANALYZERS 295–337
8.1 Introduction 295
8.2 Distortion 296
  8.2.1 Harmonic Distortion 296
  8.2.2 Intermodulation Distortion (IMD) 297
8.3 Distortion Measuring Instruments 298
  8.3.1 Distortion Analyzer 298
  8.3.2 Wave Analyzer 300
  8.3.3 Intermodulation Analyzer 300
8.4 Spectrum Analyzer 303
  8.4.1 Frequency-domain Display Capabilities 303
  8.4.2 Types of Spectrum Analyzers 304
8.5 FFT Analyzer 311
8.6 Vector Analyzer 315
8.7 Digitizing Oscilloscope 316
8.8 Logic Analyzer 318
  8.8.1 Oscilloscope or Logic Analyzer 318
  8.8.2 Types of Logic Analyzers 319
  8.8.3 Logic Timing Analyzer (LTA) 320
  8.8.4 Logic State Analyzer (LSA) 328
  8.8.5 Block Diagram of Logic Analyzer 333
  8.8.6 Interfacing Target System 334
  8.8.7 What to Look for in a Logic Analyzer? 335
Questions 336

9. ELECTRONIC COUNTERS 338–367
9.1 Introduction 338
9.2 Conventional Electronic Counters 339
  9.2.1 Basic Measurements 339
  9.2.2 Design Considerations 344
  9.2.3 Sources of Measurement Errors 350
9.3 Reciprocal Electronic Counters 354
9.4 Specifications of Electronic Counters 359
  9.4.1 Input Characteristics 359
  9.4.2 Operating Mode Specifications 360
Questions 365

Section II—Instrument Design and Manufacturing Techniques

10. GROUNDING AND SHIELDING 371–388
10.1 Introduction 371
10.2 Grounding 371
  10.2.1 The Concept of “Earth” Ground 372
  10.2.2 An Illustration of a Typical Power Supply Grounding Error 372
  10.2.3 Some Examples of Current Return Path Symbols 374

10.2.4 Shock Hazard Protection Using Earth Ground 375
10.2.5 Grounding Considerations 376
10.2.6 Basic Grounding Practices 376
10.2.7 Examples 379
10.3 Shielding 382
10.3.1 Practical Guidelines 382
10.3.2 Examples 383
10.4 Protection from Electrostatic Discharge 384
Questions *387*

11. ELEMENTS OF DESIGN 389–425
11.1 Introduction 389
11.2 Product Life Cycle 390
11.3 Circuit Design 394
11.3.1 General Issues 397
12.4 Circuit Layout 404
11.5 Assembly and Inspection 408
11.6 Testing and Calibration 408
11.7 Power Distribution 409
11.8 Wiring and Cabling 410
11.9 Enclosures 411
11.10 Integrated Testing 420
11.11 Documentation 422
Questions *424*

12. METROLOGY 426–456
12.1 Introduction 426
12.2 SI Units 428
12.3 Testing, Compatibility, Calibration, and Traceability 430
12.3.1 Testing 430
12.3.2 Compatibility 431
12.3.3 Calibration 431
12.3.4 Traceability 432
12.3.5 Measuring Equipment Cycle 434
12.3.6 Measurement Error and Measurement Uncertainty 435
12.3.7 Determining Instrument Capability 437
12.3.8 Meeting Test Specifications 438
12.3.9 Test Procedures 438
12.3.10 Reliability and Performance Metrics 439
12.3.11 National Accreditation Board of Approved Laboratories (NABAL) 440
12.3.12 Definitions of Some Commonly Used Terms 440
12.4 Electrical Standards 443
12.4.1 Voltage Standard 443
12.4.2 Resistance Standard 446

12.4.3 Current Standard 450
12.4.4 Capacitance Standard 452
12.4.5 Inductance Standard 453
12.5 Time and Frequency Standards 453
Questions *455*

13. STANDARDS IN QUALITY MANAGEMENT 457–494
13.1 Introduction 457
13.2 ISO 9000 Quality Management System 458
13.2.1 Architecture of the ISO 9000 Series of Standards 459
13.3 Quality 465
13.3.1 What is Quality? 465
13.3.2 Quality Policy 466
13.3.3 Quality Planning 466
13.3.4 Quality Management 466
13.3.5 Quality System 466
13.3.6 Quality Control 466
13.3.7 Quality Assurance 467
13.3.8 The Deming Cycle 468
13.3.9 Quality Loop 469
13.3.10 Quality Cost 469
13.3.11 Seven Wastes 471
13.3.12 Quality Improvement 471
13.3.13 Quality Audits 472
13.3.14 Quality Tools 473
13.4 Elements of ISO 9001 474
13.5 Total Quality Management (TQM) 481
13.5.1 The Steps to Total Quality Management 481
13.6 Some Other Quality Management Standards 484
13.6.1 ISO 14000 484
13.6.2 ISO 14001 484
13.6.3 ISO 17025 485
13.6.4 QS 9000 485
13.6.5 Japanese Industrial Standards (JIS) 485
13.6.6 Q-Base 486
13.6.7 E-Base 486
13.7 ISO 9000:2000 Update 487
13.7.1 Quality Management Principles 489
13.7.2 Breakdown of Sections in ISO 9000:2000 489
Questions *493*

Section III—Instrumentation Technology

14. INDUSTRIAL COMMUNICATION TECHNIQUES 497–563
14.1 Introduction 497
14.2 The OSI Network Model 500
14.3 Physical Layer Options 502
  14.3.1 Network Topologies 502
  14.3.2 Interface Standards 503
  14.3.3 RS 232 503
  14.3.4 RS 422 511
  14.3.5 RS 423 513
  14.3.6 RS 485 513
  14.3.7 IEEE 488 (GPIB) 515
  14.3.8 HART 527
  14.3.9 Token Buses and Rings 531
  14.3.10 Ethernet or CSMA/CD 533
14.4 Moving Up the Layers 535
14.5 Fieldbus and Device Networks 538
  14.5.1 The CIM Pyramid 538
  14.5.2 What is a Fieldbus? 539
  14.5.3 Why Define a Fieldbus? 540
  14.5.4 Fieldbus Requirements 546
  14.5.5 Types of Fieldbuses 547
14.6 Foundation Fieldbus (FF) 548
  14.6.1 Architecture 548
  14.6.2 Physical Layer 549
  14.6.3 Data Link Layer 554
  14.6.4 Application Layer 556
  14.6.5 FF Benefits 562
Questions 563

15. INSTRUMENTATION IN HAZARDOUS AREAS 564–596
15.1 Introduction 564
15.2 Hazardous Area Classifications 565
  15.2.1 Protective Concepts 567
15.3 Enclosure Classification Designations 569
  15.3.1 Ingress Protection (IP) Rating (or Code) 570
  15.3.2 NEMA Types 571
  15.3.3 Equipment Design and Construction 575
15.4 Intrinsically Safe Design 578
  15.4.1 Determining Safe Energy Levels 578
  15.4.2 Intrinsic Safety Circuit Design 581
  15.4.3 Installing Intrinsically Safe Systems 588
  15.4.4 Transformer Isolation Barriers (TIB) 593
15.5 Relevant Indian Standards 595
Questions 595

Appendices

A ***Attenuators*** ***599–607***
B ***Terminology Used in Hazardous Area Applications*** ***608–618***
C ***Instrumentation Symbols and P & I Diagrams*** ***619–658***
D ***Typical Specifications of Instruments*** ***659–682***
E ***Multiple Choice Questions*** ***683–719***
Answers to Multiple Choice Questions *718*

Bibliography ***721–722***

Answers to Numerical Questions ***723–724***

Index ***725–728***

Preface

There have been remarkable advancements in the design and application of electronic instruments during the last few years. Modern instruments provide digital readouts, digital data for computers and data for many other mathematical operations. However, the old type instruments are still being used at many places. There is a need to understand the operation and application of these instruments to enable engineers and scientists to adapt easily to new measurement concepts. The book presents the design features and specifications of electronic instruments and accessories.

Further, the knowledge of instrumentation techniques and technology being used in modern industries is important. The book includes topics on instrument design, industrial practices, primary standards and standards used in quality management, industrial communication techniques, etc. A separate chapter is devoted to instrumentation in hazardous environment. To draw and understand Piping and Instrumentation (P&I) diagrams is a necessary task of an instrumentation engineer. The book presents the standard instrumentation symbols and some P&I diagrams in Appendix C.

The book is divided into three sections. Section I deals with many types of the electronic instruments and accessories. The basics of the design have been presented. The meanings of the various terms used in the specifications of the instruments are also explained. Typical features of the modern instruments are given. Appendix A gives the design aspects of attenuators used in the instruments.

Good design of electronic circuits and systems begins with grounding and shielding. Generally, the engineers do not pay much attention to the grounding and shielding techniques, resulting in increase in the failure rate and electric shock hazards. Sometimes it becomes difficult to diagnose a fault due to improper grounding or shielding. Section II illustrates the procedures followed to avoid these problems.

The chapter on elements of design explains the concept of the product life cycle, general issues of circuit design, circuit layout considerations, testing and calibration methods, power distribution schemes, wiring and cabling techniques, design of enclosures, integrated testing and documentation requirements.

Metrology is the science of measurement. Various standards are used for the measurement of basic parameters. Calibration of parameters and the traceability of the laboratory standards used is an important step in maintaining the standards. Similarly, the quality management has become so important that it is now almost impossible for any industry or organization to compete without following the recommended procedures. The latest standards used, basics of testing, compatibility, calibration, traceability and various ISO 9000 Quality Assurance Guidelines are covered in Section II.

Communication methods in the industrial sector have changed so fast that communication procedures adopted earlier have become almost obsolete in some fields. Today, an instrumentation engineer needs to know the networking technology in detail. The automation has advanced to the extent of remote control and maintenance of a process from anywhere in the world. Important concepts and technology used in industrial communication are given in Section III.

Section III also enumerates requirements of hazardous areas and the interface design for instrumentation. Terminology used for hazardous areas is given in Appendix B.

Typical specifications of various instruments also form an Appendix D (courtsey different instrument manufacturers). This will help the students to know the techniques of writing the specifications and to learn about the features of modern instruments.

Problem solving is unquestionably an essential part of the learning process and the best measure of understanding. Several numerical questions are included at the end of each chapter to probe a student's grasp of the topics presented. Carefully designed qualitative multiple choice questions, chapter-wise, are given in Appendix E. Answers to numerical questions and multiple choice questions are also included.

Help of on-line documents of various manufacturers, organizations and consultants has been taken to make the book more practical and modern. This is duly acknowledged. Some of the topics are based on the lecture notes of Dr. Anil Jain, President, Vaiseshika Electronic Devices, Ambala Cantt. I owe a special debt to Dr. Anil Jain. I have been fortunate to have had the encouragement of Dr. L.K. Maheshwari, Director, BITS, Pilani and I am grateful to him for his constructive suggestions. Thanks are especially due to editorial and production staff at Prentice-Hall of India who completed the process of production, from the galley to the final printing stage, in a record time of a few days.

Last but not least, I greatly appreciate the patience and support of my family members Harmohan, Narinder, and Manpreet.

M.M.S. Anand

SECTION I

Instruments and Accessories

Chapter 1

Analog Meters

1.1 INTRODUCTION

When troubleshooting, testing, or repairing electronic equipment, we use various meters and other types of test equipment to check for proper circuit voltages, currents and resistances. The purpose of a meter, therefore, is to measure the quantities existing within a circuit. Two types of meters—analog and digital—are used. Analog meters measure the magnitude of a quantity by a proportional deflection of the meter pointer against the calibrated scale. This requires a deflection mechanism in the meter. Digital meters, on the other hand first convert the analog signal to a digital signal and then display the magnitude in digits of the quantity being measured.

Analog meters employ other circuit components as well, in addition to the basic deflection mechanism. The components may be passive or active. The meters using both passive and active components are called electronic analog meters.

In this chapter we will study the construction, operation and features of commonly used analog and electronic analog meters.

1.2 PERMANENT MAGNET MOVING COIL (PMMC) METERS

A stationary, permanent magnet moving coil (PMMC) is the basic deflection mechanism used in most analog measuring instruments. As shown in Figure 1.1(a), when no current flows through the coil, it stays in its original position. However as shown in Figure 1.1(b), when a current flows through the coil, the resulting magnetic field reacts with the magnetic field of the permanent magnet generating an electromagnetic (EM) torque, T_D, causing the movable coil to rotate. The bottom portion of the coil in Figure 1.1(a) will be the north pole of this electromagnet. Since opposite poles attract, the coil will move to the position shown in Figure 1.1(b).

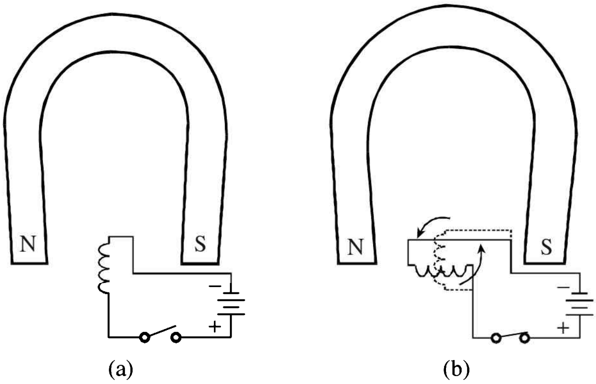


Figure 1.1 Movable coil in a magnetic field.

The relationship between the torque developed and the other parameters is given by Eq. (1.1) as

$$T_D = BANI \quad (1.1)$$

where

T_D is the developed or deflecting torque, in N-m

B is the flux density in the air gap, in Wb/m^2 (tesla)

A is the effective coil area, in m^2

N is the number of turns in the coil

I is the current in the movable coil, in amperes(A).

Typical values of the parameters are: $T_D = 3 \times 10^{-6}$ N-m, $B = 0.2$ T, $A = 1.75 \times 10^{-4}$ m^2, $N = 86$ and $I \approx 1$ mA. Coil resistance, $R_{coil} \approx 100\ \Omega$ and power dissipation $\approx 100\ \mu$W.

Since B, A and N are fixed for a given coil and magnet, T_D is directly proportional to the current through the coil, I. The coil of wire is wound on an aluminum frame or bobbin. The bobbin is supported by jeweled bearings that let it move freely (Figure 1.2).

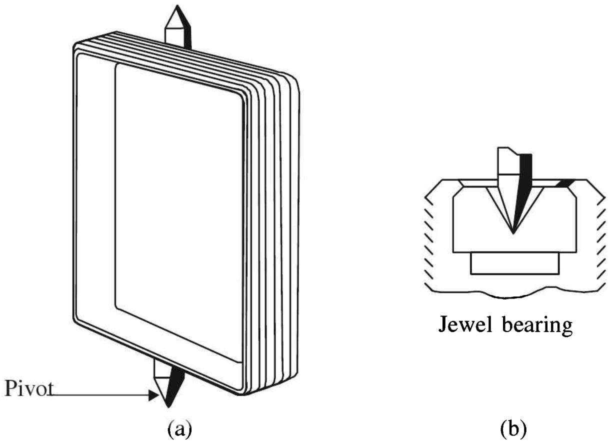


Figure 1.2 Coil arrangement: (a) bobbin and (b) jewel bearing.

To use this permanent magnet moving coil arrangement as a meter, two problems must be solved. First, a way must be found to return the coil to its original position when there is no current through the coil. Second, a method is needed to indicate the amount of coil movement.

The first problem is solved by attaching a control hairspring (generally of phosphor bronze material) to each end of the coil (Figure 1.3). These hairsprings can also be used to make the electrical connections to the coil. Hairsprings make the coil return to its initial position when there is no current. The springs also tend to resist the movement of the coil when there is current through the coil. That is, the control springs provide the opposing torque. When the developed torque exactly equals the control springs torque, the coil stops moving.

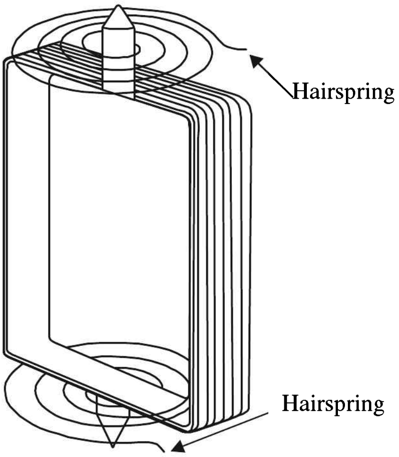


Figure 1.3 Coil and springs.

The second problem is solved by using a pointer attached to the coil which is extended out to a scale. The pointer moves as the coil moves. The scale is calibrated to indicate the amount of current through the coil (Figure 1.4).

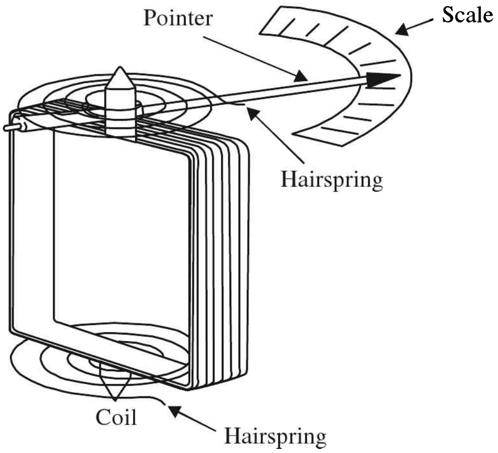


Figure 1.4 Complete coil arrangement.

Two other features are used to increase the accuracy and efficiency of this meter movement. First, an iron core is placed inside the coil to concentrate the magnetic fields. Second, curved pole pieces are attached to the magnet to ensure that the turning force on the coil increases steadily as the current increases. This is because the lines of force enter and leave a magnetic field in the air gap at right angles to the coil, regardless of the angular position of the coil. Therefore, a more linear scale is possible than if the pole faces were flat. Similar curved pole pieces are found in a motor. Figure 1.5 shows a fully assembled meter movement.

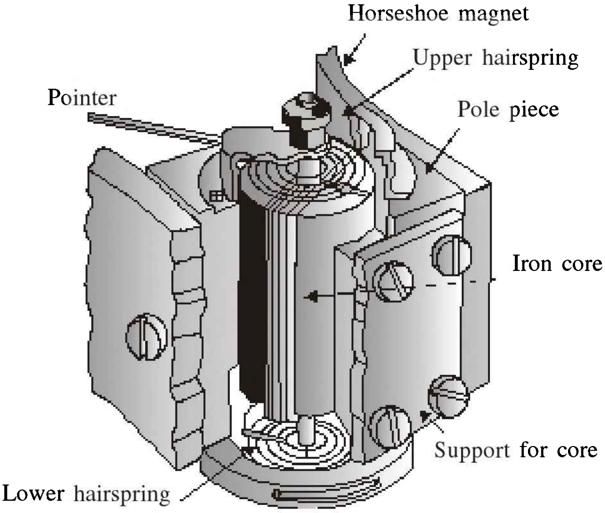


Figure 1.5 Fully assembled meter movement.

In order to ensure that the deflection is independent of the position of the instrument, it is essential that the centre of gravity of the moving system lies on the axis of rotation. To attain this balance, weights are attached to the moving system.

A permanent magnet moving coil meter mechanism is the basic mechanism for most analog measuring instruments. It is commonly called **d'Arsonval movement** because it was first employed by the Frenchman d'Arsonval in making electrical measurements. Figure 1.6 gives a view of the d'Arsonval meter movement used in a meter.

Note that if the instrument is not damped to absorb the energy of the moving element, the pointer will oscillate (vibrate) for a certain period of time before coming to a stop in its final position. Damping is an energy-absorbing system that prevents this. This is accomplished in many d'Arsonval movements by means of the motion of the aluminum bobbin on which the coil is wound. As the bobbin rotates in the magnetic field, an electromotive force is induced into it as it cuts through the lines of force. Induced currents flow in the bobbin in a direction opposite to the motion; this causes the bobbin to go beyond its final position only once before stopping. The overall sensitivity of the meter is increased by the use of a lightweight rotating assembly (bobbin, coil, and pointer) and by the use of jewel bearings.

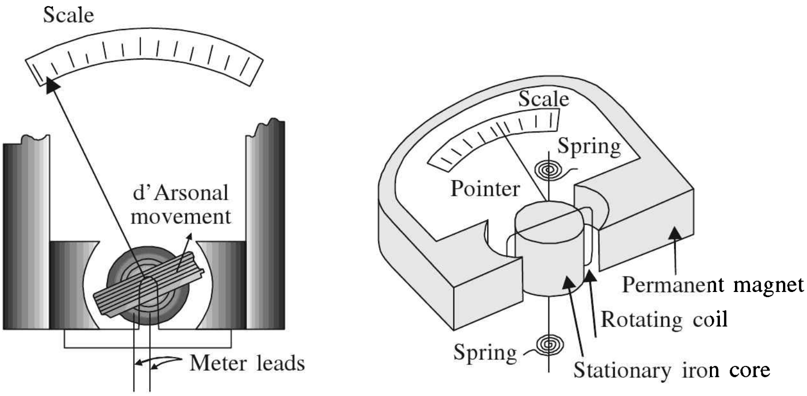


Figure 1.6 Meter using d'Arsonval movement.

Figure 1.7 shows the effect of damping on the deflection of the moving system when a step input is applied. In practice, a system is designed to be slightly underdamped to decrease its response time.

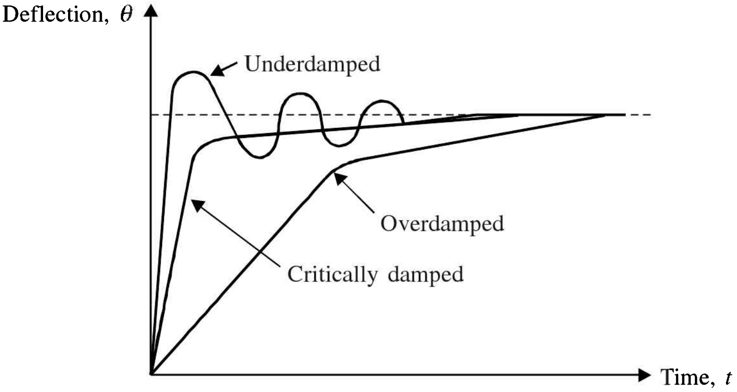


Figure 1.7 Damping curves.

Jewel bearing instruments are recommended for high vibration environments because of this mechanism's inherent stability and strength.

Taut-band suspension is another arrangement to suspend the moving mechanism, instead of using the jewel bearings. Here, the mechanism is suspended between two ribbons of metal (Figure 1.8). The absence of friction resulting from this permits greater sensitivities and provides more rugged meters. Common failures of conventional meters subjected to shock testing are cracked jewels and/or blunted pivots. For example, ANSI Specifications for panel meters require 50 G shock, whereas taut-band meters withstand 100 G shock.

Dynamic behaviour

The complete assembly of a meter forms a second-order system. The differential equation governing the motion of the coil is given by

$$J\frac{d^2\theta}{dt^2} + D\frac{d\theta}{dt} + c\theta = T_{\text{D}} = Ki(t) \tag{1.2}$$

where

J is the moment of inertia of the coil assembly
D is the damping coefficient of the system
c is the spring stiffness (or constant) of the springs
θ is the coil rotation.

The solution of Eq. (1.2) yields the response curves similar to the damping curves shown in the Figure 1.7. In most of the meters, J is quite large and hence these meters cannot follow rapid current variations. For this reason, these meters are not used for measuring variable currents. Under steady-state conditions $[(d^2\theta/dt^2) = 0,\ (d\theta/dt) = 0]$, the deflection is given by Eq. (1.3) as

$$\theta = \frac{K}{c}I \quad \text{(radians)} \tag{1.3}$$

where I is the steady-state value of the current.

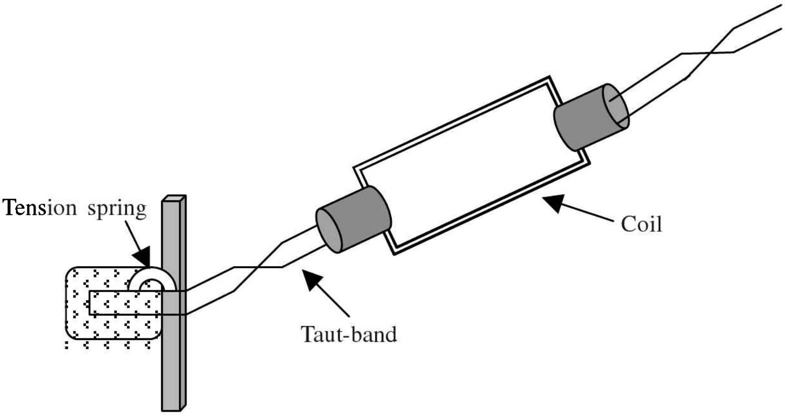


Figure 1.8 Taut-band arrangement.

Temperature compensation

The deflection of the pointer depends on the parameters which are temperature dependent. As the temperature increases, both the magnetic field strength B, and spring tension c, decrease whereas the winding resistance of the coil which is made up of copper increases. With decrease in B and increase in winding resistance, θ decreases, whereas with decrease in spring tension, θ increases. The decrease in θ due to B and increase in θ due to spring tension almost get compensated. Therefore, the temperature uncompensated meter tends to read low.

Temperature compensation is accomplished by using swamping resistors in series with the coil. The swamping resistors are made up of a material like manganin which has almost zero temperature constant. Figure 1.9 shows an arrangement for temperature compensation. The resistances R_2 and R_3 are much higher than the resistance of the coil, R_{coil}. As the temperature increases, increase in the resistance ($R_2 + R_{\text{coil}}$) is less than the increase in R_3. Therefore, more current flows through the coil. By adjusting R_1, R_2 and R_3, complete temperature compensation can be achieved. The disadvantage of using swamping resistors is the reduction in the full-scale sensitivity of the movement, since a higher applied voltage is required to achieve the full-scale deflection.

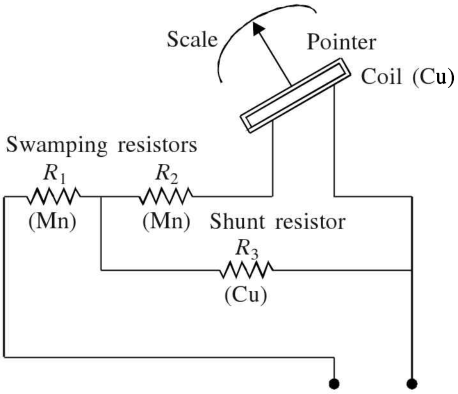


Figure 1.9 Temperature compensation arrangement.

Sources of error

The common sources of error encountered are:

- Weakening of permanent magnet due to aging
- Weakening of springs due to aging
- Change of coil resistance with temperature

These effects can be reduced by heat treatment of magnets, choosing carefully the material of springs and preaging during their manufacturing, and by proper temperature compensation techniques.

Advantages

A PMMC instrument has the following advantages:

- Uniform scale
- Not affected by stray magnets
- Low power consumption
- Eddy current damping is used which is most effective
- No hysteresis loss

Disadvantages

The disadvantages of a PMMC instrument are:

- Can be used for dc measurements only
- Expensive
- Aging effect on springs used for controlling

Circuit representation

Figure 1.10 shows the circuit representation of the PMMC movement. I_m is the full-scale deflection current through the mechanism and R_m is the resistance of the mechanism.

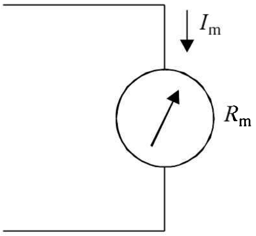


Figure 1.10 Circuit representation of PMMC deflection mechanism.

1.3 DC ANALOG METERS

1.3.1 DC Ammeter

A meter designed to measure electrical current is called *ammeter*, because the unit of measurement is in amperes. PMMC movement as such can be used to measure direct current. But since this movement can carry only very small currents, for large current measurements, a resistance is connected in parallel to the movement to bypass additional current. This resistance is called a *shunt* and is denoted as R_{sh}. The basic dc ammeter is shown in Figure 1.11. Out of the input current I, current equal to I_m passes through the meter and the balance current $I_{sh} = I - I_m$ flows through the shunt resistance (as per Kirchhoff's current law, KCL).

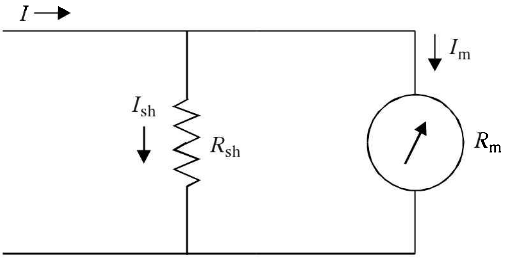


Figure 1.11 Basic dc ammeter circuit.

We can calculate the value of R_{sh} by using Kirchhoff's voltage law (KVL), according to which voltage across R_m and R_{sh} are equal. Therefore,

$$I_{sh} R_{sh} = I_m R_m$$

or

$$R_{sh} = \frac{I_m R_m}{I_{sh}} = \frac{I_m R_m}{I - I_m} = \frac{R_m}{n - 1}$$

where $n = I/I_m$.

Example 1.1

A 1 mA meter movement with an internal resistance of 50 Ω is to be used in a 0 to 100 mA ammeter. Find the value of the required shunt resistance.

Solution

$$n = \frac{I}{I_m} = \frac{100 \text{ mA}}{1 \text{ mA}} = 100$$

$$R_{sh} = \frac{R_m}{n - 1} = \frac{50\ \Omega}{100 - 1} = 0.5050\ \Omega$$

Internal and external shunts

The shunt resistance, R_{sh}, is normally a wire of very low temperature coefficient. It may be connected within the case of the meter (internal shunt) or external to the meter (external shunt). The value of the shunt is very low for large input currents. Special shunts are prepared and are generally connected outside due to their large size and heating. Figure 1.12 shows a typical external shunt.

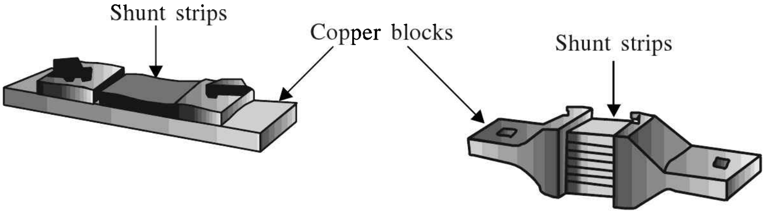


Figure 1.12 Typical ammeter shunts.

The shunt strips are usually made of manganin which has a temperature coefficient of almost zero. The zero-temperature coefficient property is desirable because of the large currents that often flow through shunts, producing heat. A zero-temperature coefficient material is not affected by this heat; therefore, it remains stable at high temperatures. The resistance of most other materials increases on being heated. If shunts were made of these materials, they would begin to carry less current with increase in their temperature. More and more current would then flow through the meter movement, increasing the chances of damage. Using shunts constructed with zero-temperature coefficient materials eliminates this problem.

The ends of the shunt strips are embedded in heavy copper blocks. The blocks are attached to the meter coil leads and the line terminals. To ensure accurate readings, we should not interchangeably use the meter leads for a particular ammeter with those for a meter of a different range. Slight changes in lead length and size may vary the resistance of the meter circuit. If this happens, current will also change and produce incorrect meter readings. External shunts are generally used where currents greater than 50 amperes must be measured.

Multirange ammeter

Various values of shunt resistances can be used by means of a suitable switching arrangement, to increase the number of current ranges that can be covered by the meter. Two switching arrangements are shown in Figure 1.13. Figure 1.13(a) is a simpler arrangement when the position of the switch S, connects one of the shunts to the meter. Figure 1.13(b) displays the Ayrton's shunt arrangement, which is also called *universal shunt.*

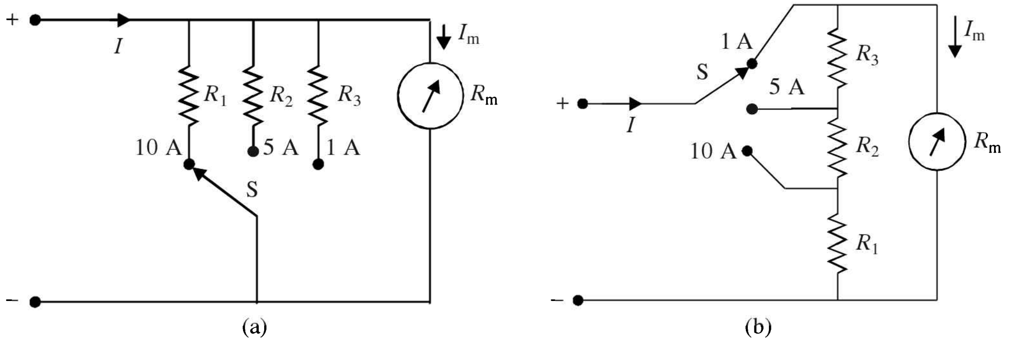


Figure 1.13 Ways to connect internal shunts: (a) simple arrangement and (b) Ayrton shunt arrangement.

Example 1.2

A 1 mA meter deflecting mechanism with an internal resistance of 50 Ω is to be used in a 0 to 1 A, 0 to 5 A and 0 to 10 A ranges in the arrangement shown in Figure 1.13(a). Find the value of the required shunt resistances.

Solution

0 to 1 A range

$$n = \frac{I}{I_m} = \frac{1\,\text{A}}{1\,\text{mA}} = 1000$$

$$R_3 = \frac{R_m}{n-1} = \frac{50\ \Omega}{1000-1} = 0.05005\ \Omega$$

0 to 5 A range

$$n = \frac{I}{I_m} = \frac{5\,\text{A}}{1\,\text{mA}} = 5000$$

$$R_2 = \frac{R_m}{n-1} = \frac{50\ \Omega}{5000-1} = 0.01002\ \Omega$$

0 to 10 A range

$$n = \frac{I}{I_m} = \frac{10\,\text{A}}{1\,\text{mA}} = 1000$$

$$R_1 = \frac{R_m}{n-1} = \frac{50\ \Omega}{10000-1} = 0.00500\ \Omega$$

The arrangement shown in Figure 1.13(a) has two disadvantages:

1. When the switch is moved from one shunt resistor to another, the shunt is momentarily removed from the meter. The full input current then flows through the meter coil. Even a momentary surge of current can easily damage the coil.

2. The contact resistance of the switch is in series with the shunt, but not with the meter coil. In shunts that must pass high currents, this contact resistance becomes an appreciable part of the total shunt resistance. Because the contact resistance is of a variable nature, the ammeter indication may not be accurate.

The arrangement of Figure 1.13(b) eliminates the possibility of the meter being momentarily disconnected from the shunt. Further, the contact resistance of the range-selector switch is external to the shunt and meter in each range position. The contact resistance in this case therefore has no effect on the accuracy of the measurement of current.

Example 1.3

A 1 mA meter movement with an internal resistance of 50 Ω is to be used in a 0 to 1 A, 0 to 5 A and 0 to 10 A ranges in the arrangement shown in Figure 1.13(b). Find the value of the required shunt resistances.

Solution

0 to 1 A range

$$n = \frac{I}{I_m} = \frac{1\,\text{A}}{1\,\text{mA}} = 1000$$

$$R_1 + R_2 + R_3 = \frac{R_m}{n-1} = \frac{50\ \Omega}{1000-1} = 0.05005\ \Omega$$

0 to 5 A range

$$n = \frac{I}{I_m} = \frac{5\,\text{A}}{1\,\text{mA}} = 5000$$

$$R_1 + R_2 = \frac{R_m + R_3}{n - 1} = \frac{50\ \Omega + R_3}{5000 - 1}$$

0 to 10 A range

$$n = \frac{I}{I_m} = \frac{10\,\text{A}}{1\,\text{mA}} = 1000$$

$$R_1 = \frac{R_m + R_2 + R_3}{n - 1} = \frac{50\ \Omega + R_2 + R_3}{10000 - 1}$$

Solving above equations simultaneously, we get

$$R_1 = 0.005005\ \Omega, \quad R_2 = 0.005005\ \Omega, \quad R_3 = 0.04004\ \Omega$$

Note that for 0 to 5 A range, R_3 is in series with R_m, and for 0 to 10 A range R_2, R_3 and R_m are in series. Therefore it can be seen that for higher current ranges, the overall meter resistances increase slightly.

Connecting an ammeter in a circuit

Ammeters must always be connected *in series* with a circuit, never in parallel with it. If an ammeter is connected across a constant-voltage source of appreciable voltage, the low internal resistance of the meter will bypass the normal circuit resistance. This results in the application of source voltage (or a good portion of it) directly to meter terminals. The resulting excessive current will burn up the meter coil and render the meter useless.

If we do not know the approximate value of current in the circuit, we should take a reading at the highest range of the ammeter; then we should switch progressively to lower ranges to obtain a suitable reading. Most ammeter scales indicate the current being measured in increasing values from left to right. If we connect the meter without observing proper polarity, the pointer will be deflected backwards (from right to left) which may damage the meter movement.

Ammeter sensitivity

Ammeter sensitivity is determined by the amount of current required by the meter coil to produce full-scale deflection of the pointer. The smaller the amount of current required to produce this deflection, the greater the sensitivity of the meter. A meter movement that requires only 100 μA for full-scale deflection has a greater sensitivity than a meter movement that requires 1 mA for the same deflection.

Good sensitivity is especially important in ammeters to be used in circuits in which small currents flow. As the meter is connected in series with the load, the load current flows through the meter. If the internal resistance of the meter is a large portion of the load resistance, an effect known as *meter-loading* occurs. Meter-loading is the condition that occurs when the insertion of a meter into a circuit changes the operation of that circuit. We then say that the meter is loading the circuit. This condition is not desirable. The purpose of inserting a meter into a circuit is to allow the measurement of circuit current under the normal operating conditions. If the meter changes the circuit operation and changes the amount of current flow, the reading obtained will have an error. The loading effect is shown in Figure 1.14.

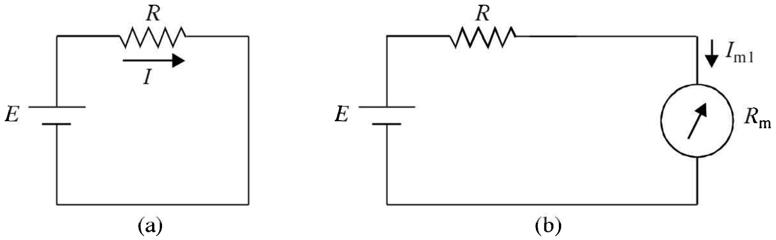


Figure 1.14 Ammeter loading effect.

The current, I, in the circuit of Figure 1.14(a) is given by the equation

$$I = \frac{E}{R} \tag{1.4}$$

When the ammeter is connected in the circuit (Figure 1.14(b)), the current read by the circuit, I_{m1}, is given by the equation

$$I_{m1} = \frac{E}{R + R_m} \tag{1.5}$$

Dividing Eq. (1.5) by Eq. (1.4),

$$\frac{I_{m1}}{I} = \frac{R}{R + R_m}$$

Ideally, $I_{m1}/I = 1$, i.e. the meter should read the same current and should not alter the current. But due to loading, $I_{m1} < I$. The lower the value of R_m compared with R, the lower the loading effect would be, i.e. $I_{m1}/I \rightarrow 1$.

Example 1.4

(a) A 1 mA meter movement with an internal resistance of 50 Ω is connected in series in the circuit shown in Figure 1.14(a). The values of E and R are 50 mV and 50 Ω respectively. Find the loading effect of the meter on the circuit.

(b) If the resistance of the meter is changed to 5 Ω, calculate the revised loading effect.

Solution

(a) Before connecting the meter

$$I = \frac{50 \text{ mV}}{50\ \Omega} = 1 \text{ mA}$$

After connecting the meter

$$I_{m1} = \frac{50 \text{ mV}}{100\ \Omega} = 0.5 \text{ mA}$$

$$\frac{I_{m1}}{I} = 0.5, \text{ i.e. } 50\% \text{ error}$$

(b) In this case, I_{m1} = 0.909 mA. The error is therefore equal to 9.1% only.

Therefore, it is proved that the lower the resistance of the meter, the lower the loading effect would be.

Ammeter safety precautions

When using an ammeter, certain precautions must be observed to prevent injury to personnel and damage to the ammeter or the equipment, or the circuit in which the ammeter is connected. The following safety precautions should be observed while using an ammeter:

- Always connect an ammeter in series with the circuit under test.
- Always start with the highest range on an ammeter (or any meter). Be aware of the loading effect caused by the ammeter on the circuit.
- De-energize and discharge the circuit completely before connecting or disconnecting the ammeter.
- Observe the proper circuit polarity to prevent the meter from being damaged.
- Never use a dc ammeter to measure alternating current.
- Observe the general safety precautions of electricity.

1.3.2 DC Voltmeter

The d'Arsonval mechanisms have full-scale deflection current ratings as little as 50 μA, with an internal resistance of less than 1000 Ω. This means we can connect a voltage source of only 50 mV (50 μA × 1000 Ω) across the mechanism. That is, a voltage of only 50 mV can be measured by this mechanism. In order to build voltmeters with practical (higher voltage) scales from such sensitive movements, we need to find some way to reduce the measured quantity of voltage down to a level that the movement can handle.

We need a voltage divider circuit to proportion the total input voltage in such a way that the voltage across the movement does not cross its ratings. A resistor, R_s, is connected in series with the meter movement. Figure 1.15 shows the arrangement.

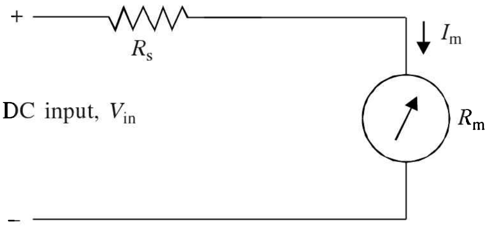


Figure 1.15 Basic dc voltmeter circuit.

The series resistor is called a *multiplier* resistor since it multiplies the working range of the meter movement as it proportionately divides the measured voltage across it. To determine multiplier resistance, R_s, we apply KVL across the loop. Thus,

$$V_{in} = I_m(R_s + R_m) \quad \text{or} \quad R_s = \frac{V_{in}}{I_m} - R_m$$

Since in voltmeters we measure voltage, the scale is calibrated in volts.

Multirange voltmeter

Multirange voltmeters use a common meter movement. The required resistances are connected in series with the meter by a switching arrangement. A schematic diagram of a multirange voltmeter with four ranges is shown in Figure 1.16.

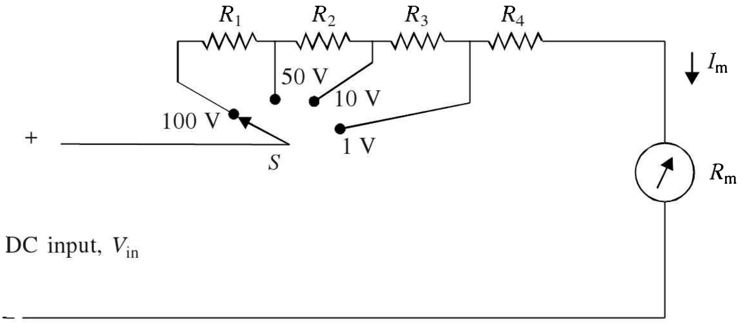


Figure 1.16 Multirange voltmeter arrangement.

To calculate the values of the resistors, let us consider Example 1.5.

Example 1.5

A 1 mA meter movement with an internal resistance of 50 Ω is to be used in a 0–1 V, 0–10 V, 0–50 V, and 0–100 V ranges in the arrangement shown in Figure 1.16. Find the value of the required multiplier resistances.

Solution

0 to 1 V range

$$R_4 = \frac{V_{in}}{I_m} - R_m = \frac{1}{1 \times 10^{-3}} - 50 = 950\ \Omega$$

0 to 10 V range

$$R_3 = \frac{V_{in}}{I_m} - (R_m + R_4) = \frac{10}{1 \times 10^{-3}} - (50 + 950) = 9\ \text{k}\Omega$$

0 to 50 V range

$$R_2 = \frac{V_{in}}{I_m} - (R_m + R_4 + R_3) = \frac{50}{1 \times 10^{-3}} - (50 + 950 + 9000) = 40\ \text{k}\Omega$$

0 to 100 V range

$$R_1 = \frac{V_{\text{in}}}{I_{\text{m}}} - (R_{\text{m}} + R_4 + R_3 + R_2) = \frac{100}{1 \times 10^{-3}} - (50 + 950 + 9000 + 40000) = 50\ \text{k}\Omega$$

Note that resistance offered by the meter at the input terminals in each range is 1 kΩ per volt. Further, except R_4, all other values of the resistors are readily available.

Connecting a voltmeter in a circuit

A voltmeter must always be connected in *parallel* with a circuit. If we are not sure of the level of the voltage to be measured, we can take a reading at the highest range of the voltmeter and progressively (step-by-step) lower the range until a suitable reading is obtained. Observing the correct polarity is important when connecting an instrument in a circuit.

Voltmeter sensitivity (ohm per volt rating)

The sensitivity of a voltmeter is given in ohms per volt. It is determined by dividing the sum of the resistance of the meter (R_{m}) plus the series resistance (R_{s}), by the full-scale reading in volts. In equation form, sensitivity is expressed as follows:

$$\text{Sensitivity, } S = \frac{R_{\text{m}} + R_{\text{s}}}{V_{\text{in}}}$$

This is same as saying that the sensitivity is equal to the reciprocal of the full-scale deflection current. In equation form, this is expressed as follows:

$$\text{Sensitivity, } S = \frac{\text{Ohm}}{\text{Volt}} = \frac{1}{\text{Amp}} = \frac{1}{I_{\text{fsd}}}$$

Thus, the sensitivity of a 1 mA movement is the reciprocal of 1 mA, or 1 kΩ per volt. This we have seen in Example 1.5.

Loading effect

The purpose of a voltmeter is to indicate the potential difference between two points in a circuit. When a voltmeter is connected across a circuit, it shunts the circuit. If the voltmeter has a low resistance, it draws a substantial amount of current. This action lowers the effective resistance of the circuit and changes the voltage reading. This is called the *loading effect.*

Loading effect is caused due to low sensitivity of the voltmeter. When we make voltage measurements in high-resistance circuits, we should use a high sensitivity voltmeter to prevent the loading effect of the voltmeter. The effect is less noticeable in low-resistance circuits because the loading effect is less. Examples 1.6 and 1.7 illustrate the loading effect on measurements.

Example 1.6

In the circuit shown in Figure 1.17, the voltage across the resistor of value 25 kΩ is to be measured first by using a voltmeter of sensitivity of 1 kΩ/V, and then with a voltmeter of sensitivity of 20 kΩ/V. Calculate the reading of the voltmeter in each case and the % error in the measurement.

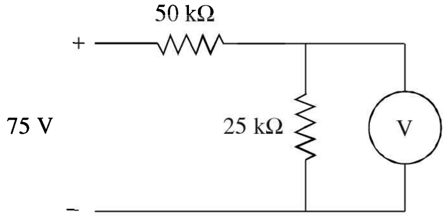


Figure 1.17 Voltmeter loading effect.

Solution

True value of the voltage across 25 kΩ resistor = 25 V. Therefore, the voltmeter should read this voltage. To get maximum accuracy in measurement, let us choose the voltmeter range of 25 V.

For S = 1 kΩ/V

Voltmeter resistance in 25 V range = 25 × 1 kΩ = 25 kΩ

$$\text{Voltage measured by the voltmeter} = \frac{25\text{ k}\Omega \,||\, 25\text{ k}\Omega}{50\text{ k}\Omega + (25\text{ k}\Omega \,||\, 25\text{ k}\Omega)} \times 75 = 15\text{ V}$$

$$\%\text{ error} = \frac{25 - 15}{25} \times 100 = 40\%$$

For S = 20 kΩ/V

Voltmeter resistance in 25 V range = 25 × 20 kΩ = 500 kΩ

$$\text{Voltage measured by the voltmeter} = \frac{25\text{ k}\Omega \,||\, 500\text{ k}\Omega}{50\text{ k}\Omega + (25\text{ k}\Omega \,||\, 500\text{ k}\Omega)} \times 75 = 24.2\text{ V}$$

$$\%\text{ error} = \frac{25 - 24.2}{25} \times 100 = 3.2\%$$

Example 1.7

If the resistors in the circuit shown in Figure 1.17 are changed to 5 kΩ and 2.5 kΩ from 50 kΩ and 25 kΩ, calculate the reading of the voltmeter with sensitivity of 1 kΩ/V and find the % error in the measurement.

Solution

True value of the voltage across 2.5 kΩ resistor is still the same, i.e. 25 V. To get maximum accuracy in measurement, let us again choose the voltmeter range of 25 V.

Voltmeter resistance in 25 V range = 25 × 1 kΩ = 25 kΩ

$$\text{Voltage measured by the voltmeter} = \frac{2.5\ \text{k}\Omega \,||\, 25\ \text{k}\Omega}{5.0\ \text{k}\Omega + (2.5\ \text{k}\Omega \,||\, 25\ \text{k}\Omega)} \times 75 = 23.44\ \text{V}$$

$$\%\ \text{error} = \frac{25 - 23.44}{25} \times 100 = 6.24\%$$

From Examples 1.6 and 1.7, we observe that the voltmeter with the same sensitivity gives a large error in one case and a small error in the other. The loading effect is less in the circuit with low resistance values. But the use of low resistance values in the circuit, increases the current drain from the source.

Voltmeter safety precautions

Just as in case of ammeters, safety precautions need to be observed with voltmeters too, to prevent injury to personnel and damage to the voltmeter or equipment. The following safety precautions should be observed while using a voltmeter:

- Always connect a voltmeter across a circuit.
- Always begin with the highest range of a voltmeter. Be aware of the loading effect.
- In dc voltmeters, observe the proper circuit polarity to prevent damage to the meter.
- Never use a dc voltmeter to measure ac voltage.
- Observe the general safety precautions of electricity.

1.3.3 Ohmmeter

An ohmmeter is an instrument used to measure resistance and check the continuity of electrical circuits and components. This resistance reading is indicated through a meter movement. The ohmmeter must then have an internal source of voltage to create the necessary current to operate the movement, and also have appropriate ranging resistors to allow desired current to flow through the movement at any given resistance.

Two types of schemes are used to design an ohmmeter—*series type* and *shunt type*. The series type of ohmmeter is used for measuring relatively high values of resistances, while the shunt type is used for measuring low values of the resistances.

Series type ohmmeter

Figure 1.18 shows the basic circuit of the series type ohmmeter.

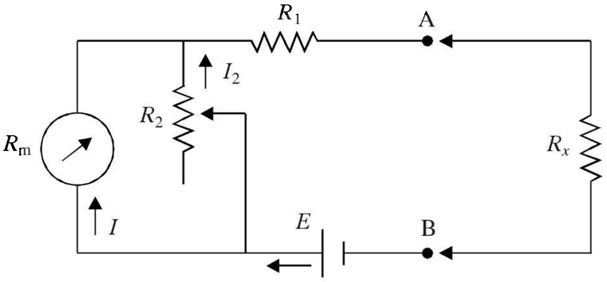


Figure 1.18 Basic series type ohmmeter.

In this figure, R_1 is the current limiting resistor, R_2 is the zero adjust resistor, R_x is the unknown resistor, E is the internal battery voltage and R_m is the internal resistance of the d'Arsonval movement. A and B are the output terminals of the ohmmeter across which an unknown resistor is connected.

When $R_x = 0$ (short circuit), R_2 is adjusted to get full-scale current through the movement. Then, $I = I_{fsd}$. The pointer will be deflected to its maximum position on the scale. Therefore, this full-scale current reading is marked 0 ohms.

When $R_x = \infty$ (open circuit), $I = 0$. The pointer will read zero. Therefore, the zero current reading is marked ∞ ohms.

By connecting different values of R_x, intermediate values are marked. The overall accuracy of the scale markings depends on the repeating accuracy of the movement and tolerances of the resistors used for calibration. Figure 1.19 shows a typical scale of the series type ohmmeter. Note that the scale is logarithmic— "expanded" at the low end of the scale and "compressed" at the high end to be able to span a wide range from zero to infinite resistance.

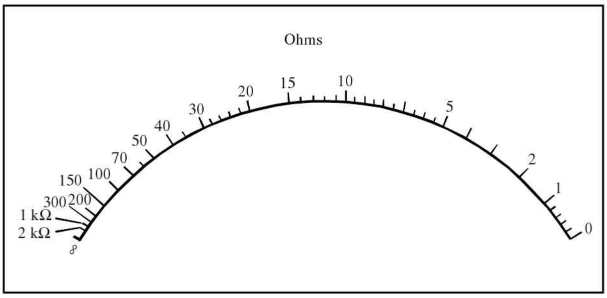


Figure 1.19 Typical scale of series type ohmmeter.

To calculate R_1 and R_2

R_1 and R_2 used in Figure 1.18 can be calculated by using a value of R_x corresponding to half the deflection of the meter. For the given movement, I_{fsd} and R_m are known.

Let R_h be the half deflection resistance. For this value of R_x, $I = I_{fsd}/2$.

Further, at half deflection,

R_h = internal resistance of the circuit looking from terminals A and B.

$$= R_1 + \frac{R_2 R_m}{R_2 + R_m}$$

Battery current needed to supply half-scale deflection is given by

$$I_h = \frac{E}{2R_h}$$

Total current, I_t, supplied by the battery for full-scale deflection is double of this current, i.e.

$$I_t = \frac{E}{R_h}$$

and

$$I_2 = I_t - I_{fsd}$$

Using KVL

$$I_2 R_2 = I_{fsd} R_m$$

Solving these equations, we get

$$R_2 = \frac{I_{fsd} R_m R_h}{E - I_{fsd} R_h}; \qquad R_1 = R_h - \frac{I_{fsd} R_m R_h}{E}$$

Example 1.8

In the circuit of Figure 1.18, a 1 mA meter movement with an internal resistance of 50 Ω is to be used. The battery voltage is 3 V. Half-scale deflection should be for 2500 Ω.

(a) Calculate the values of R_1 and R_2.
(b) Find the change in the value of R_2 if the battery voltage reduces by 10%.
(c) What is the half-scale deflection if battery voltage reduces by 10%?

Solution

(a) $$R_2 = \frac{I_{fsd} R_m R_h}{E - I_{fsd} R_h} = \frac{10^{-3} \times 50 \times 2500}{3 - 10^{-3} \times 2500} = 250\ \Omega$$

$$R_1 = R_h - \frac{I_{fsd} R_m R_h}{E} = 2500 - \frac{10^{-3} \times 50 \times 2500}{3} = 2458.33\ \Omega$$

(b) $$E = 3 - 0.3 = 2.7\ \text{V}$$

$$R_2 = \frac{10^{-3} \times 50 \times 2500}{2.7 - 10^{-3} \times 2500} = 625\ \Omega$$

R_2 should be changed from 250 Ω to 625 Ω to compensate for this reduction in battery voltage.

(c) $$R_h = R_1 + \frac{R_2 R_m}{R_2 + R_m} = 2458.33 + \frac{625 \times 50}{675} = 2504.63\ \Omega$$

Half-scale deflection now corresponds to 2504.63 Ω instead of 2500 Ω.

Shunt-type ohmmeter

Figure 1.20 shows the basic circuit of the shunt-type ohmmeter where movement mechanism is connected parallel to the unknown resistance. In this circuit it is necessary to use a switch, otherwise current will always flow in the movement mechanism. Resistor R_{sh} is used to

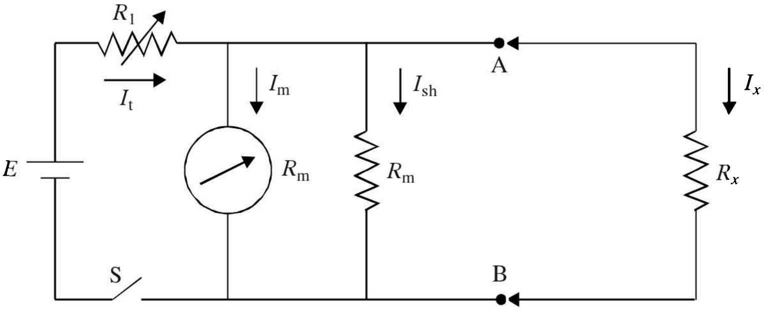


Figure 1.20 Basic shunt-type ohmmeter.

bypass excess current. Let the switch be closed. When $R_x = 0$ (short circuit), the pointer reads zero because full current flows through R_x and no current flows through the meter and R_{sh}. Therefore, zero current reading is marked 0 ohms. When $R_x = \infty$ (open circuit), no current flows through R_x. Resistor R_1 is adjusted so that full-scale current flows through the meter. Therefore, maximum current reading is marked ∞ ohms. Comparison of series and shunt ohmmeter scales is shown in Figure 1.21.

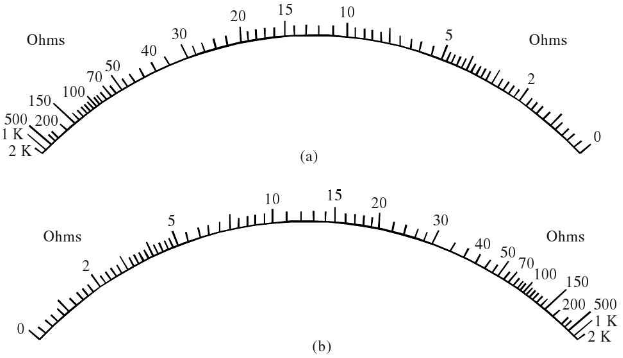


Figure 1.21 Ohmmeter scales: (a) series scale and (b) shunt scale.

To calculate R_1 and R_{sh}. Again, we can use the concept of half-scale deflection. Let R_h be the half-deflection resistance. For this value of R_x, $I_m = I_{fsd}/2$.

Further, at half deflection, current through R_h is equal to sum of the currents through R_{sh} and R_m, i.e.

$$I_x = I_{sh} + I_m$$

Also

$$I_x = \frac{E}{R_h} = \frac{I_m R_m}{R_h}$$

Solving for I_{sh}, we get

$$I_{sh} = I_m \left(\frac{R_m - R_h}{R_h} \right)$$

Therefore,

$$R_{sh} = \frac{I_m R_m}{I_{sh}}$$

Now,

$$I_t = I_m + I_{sh} + I_x = 2(I_m + I_{sh})$$

$$= 2I_m\left(\frac{R_m}{R_h}\right)$$

Therefore,

$$R_1 = \frac{E - I_m R_m}{I_t}$$

Example 1.9

In the circuit of Figure 1.20, a 1 mA meter movement with an internal resistance of 50 Ω is to be used. The battery voltage is 3 V. Half-scale deflection should be for 0.5 Ω. Calculate the values of R_1 and R_{sh}.

Solution

For half-scale deflection

$$I_m = 0.5 I_{fsd} = 0.5 \text{ mA}$$

$$I_{sh} = I_m\left(\frac{R_m - R_h}{R_h}\right) = 0.5 \times 10^{-3}\left(\frac{50 - 0.5}{0.5}\right) = 45 \text{ mA}$$

Therefore,

$$R_{sh} = \frac{I_m R_m}{I_{sh}} = 0.5 \times \frac{50}{45} = \frac{5}{9}\ \Omega$$

$$I_t = 2I_m\left(\frac{R_m}{R_h}\right) = 2 \times 0.5 \times 10^{-3} \times \frac{50}{0.5} = 100 \text{ mA}$$

$$R_1 = \frac{E - I_m R_m}{I_t} = \frac{3 - 0.5 \times 10^{-3} \times 50}{100 \times 10^{-3}}$$

$$= 29.75\ \Omega$$

We can thus see that shunt-type ohmmeter can measure low values of resistance.

Ohmmeter safety precautions

The following safety precautions and operating procedures while using ohmmeters should be observed:

- Be certain that the circuit is de-energized and discharged before connecting an ohmmeter.

- Do not apply power to the circuit while measuring resistance.
- When finished using the ohmmeter, switch it to the OFF position.
- Always adjust the ohmmeter for zero reading after changing a range and before making resistance measurement.

1.3.4 Multimeter

A multimeter is the most common measuring instrument. The name multimeter comes from **multiple meter**. It is a dc ammeter, a dc voltmeter, an ac voltmeter, and an ohmmeter, all in one package. Figure 1.22 is a picture of a typical multimeter.

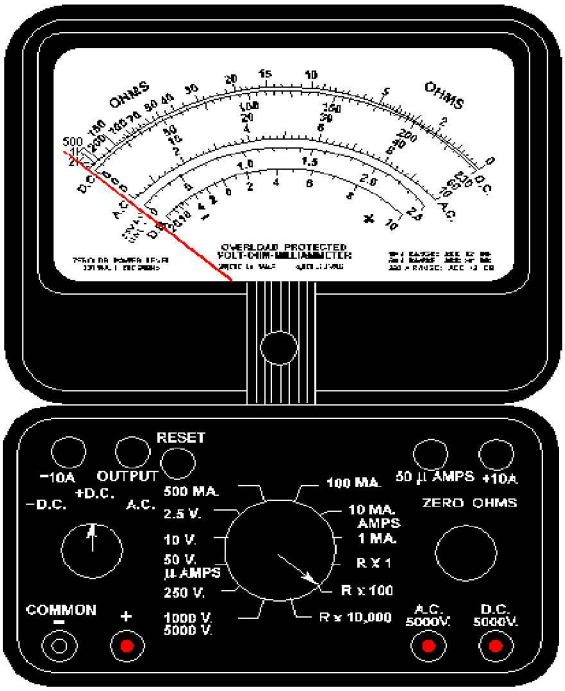


Figure 1.22 A typical multimeter.

Most multimeters use a d'Arsonval meter movement and have a built-in rectifier for ac measurements. The lower portion of the meter shown in Figure 1.22 contains the function switches and jacks (for the meter leads). Separate jacks are provided for high voltage and high current measurements. The ac voltage scale reads the rms voltage for the sinusoidal input.

The function switch is used to select the type and range of the measurement to be performed. As is obvious, a common d'Arsonval meter movement is used along with the circuits used for dc ammeter (Figure 1.13(a)), dc voltmeter (Figure 1.16), and ohmmeter (Figure 1.18). For resistance measurements, generally, three multiplication ranges ($R \times 1$,

$R \times 10$, $R \times 100$) are used. Each range selects a different value of the resistor, R_1 in Figure 1.18. R_1 is maximum for $R \times 1$ range and minimum for $R \times 100$ range. For high values of resistance measurements, the internal battery voltage should be increased. Most of the multimeters use a 9 V battery.

Note that the divisions on the scale are easier to read on the right-hand portion of the scale than those on the left. For this reason, if an unknown resistance is greater than, say, 3 kΩ and if we are using the $R \times 1$ range, we will be unable to accurately read the indicated resistance. If we change the range switch to the $R \times 10$ position, the pointer will now deflect to a reading of 300 Ω. This indicates that the unknown resistance has 3 kΩ (300 times 10) of resistance. If the $R \times 100$ range were to be used to measure the same 3 kΩ resistor, the pointer would deflect still further to the 30 Ω position.

Multimeters often provide a dBm scale. Let us find out more about this scale.

Decibel (dBm) scale

Measurements in audio frequency are often expressed in decibels since the wide range of frequencies cannot be accommodated on a linear scale. The decibel is a logarithmic unit which compresses this wide range down to a more easily handled range. Order-of-magnitude (factor-of-10) changes the results in equal increments on a decibel scale. Furthermore, the human ear perceives changes in amplitude on a logarithmic basis, making measurements with the decibel scale reflect audibility more accurately.

A decibel may be defined as the logarithmic ratio of two power measurements or as the logarithmic ratio of two voltages. Following equations define the decibel for both power and voltage measurements.

$$\text{dB} = 10 \log \frac{P_1}{P_2}$$

$$\text{dB} = 20 \log \frac{V_1}{V_2}$$

There is no difference between the decibel values from power measurements and the decibel values from voltage measurements if the impedances are equal. In both equations the denominator variable is usually a stated reference. Whether the decibel value is computed from the power-based equation or from the voltage-based equation, the same result is obtained. A doubling of voltage yields a value of 6.02 dB, while a doubling of power yields 3.01 dB. This is true because doubling voltage results in a factor-of-4 increase in power.

Often the decibel value of a signal is expressed relative to some standard reference instead of another signal. The reference for decibel measurements may be predefined as a power level, as in dBm (decibels above 1 mW), or it may be a voltage reference. When measuring dBm or any power-based decibel value, the reference impedance must be specified or understood. For example, 0 dBm (600 Ω) is the correct way to specify the level. 600 Ω is a common reference impedance in audio circuits and 50 Ω in microwave circuits. The equations assume that the circuit being measured is terminated in the reference impedance used in the decibel calculation. However, most voltmeters are high-impedance devices and are calibrated in decibels relative to the voltage required to reach 1 mW in the reference impedance. This voltage is 0.775 V in the 600 Ω. Termination of the line in 600 Ω is left to the user. If the line is not terminated, it is not correct to speak of a dBm measurement. The case of decibels in an unloaded line is referred to as dBu (or sometimes dBv) to denote that it is referenced to a 0.775 V level without regard to impedance.

Another common decibel reference used in voltage measurements is 1 V. When using this reference, measurements are presented as dBV. Often it is desirable to specify levels in terms of a reference transmission level somewhere in the system under test. These measurements are designated dBr where the reference point or level must be separately conveyed.

Multimeter safety precautions

The following safety precautions and operating procedures for ohmmeters should be observed:

- De-energize and discharge the circuit completely before connecting or disconnecting a multimeter.
- Never apply power to the circuit while measuring resistance with a multimeter.
- Connect the multimeter in series with the circuit for current measurements, and in parallel for voltage measurements.
- Be certain the multimeter is switched to ac before attempting to measure ac circuits.
- Observe the proper dc polarity when measuring dc values.
- When finished with a multimeter use, switch it to the OFF position, if available. If there is no OFF position, switch the multimeter to the highest ac voltage position.
- Always start with the highest voltage or current range.
- Select a final range that allows a reading near the middle of the scale.
- Adjust the “0 ohms” reading after changing resistance ranges and before making a resistance measurement.
- Be certain to read ac measurements on the ac scale of a multimeter.
- Observe the general safety precautions for electrical and electronic devices.

1.3.5 Calibration

Calibration is necessary to ensure the accuracy of an instrument against the national and international standards. Figure 1.23 shows the circuits for the calibration of dc ammeters and dc voltmeters.

As shown in the Figure 1.23(a), the voltage across a standard resistor is measured by a standard instrument (voltmeter or potentiometer), current is calculated by dividing this voltage with the value of standard resistor which is then compared with the current shown by the ammeter under test. If both the currents are not equal, ammeter is set to read the calculated value of the current. The method is repeated for different current values obtained by varying the rheostat resistance.

Figure 1.23(b) shows the setup for voltmeter calibration. Voltages measured by the standard instrument and the voltmeter under test are compared. If both the voltages are not

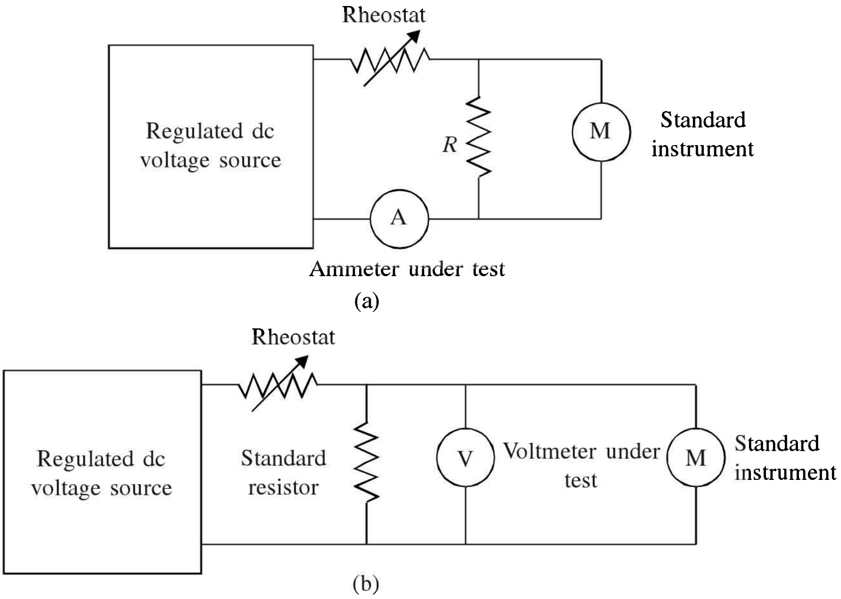


Figure 1.23 Calibration of (a) dc ammeter and (b) dc voltmeter.

equal, the voltmeter is set to read the standard value of the voltage. The method is repeated for different voltage values obtained by varying the rheostat resistance.

The ohmmeter circuit is considered to be of moderate accuracy and low precision. Its rough calibration can be performed by measuring values of standard resistors with it and recording the readings obtained.

1.4 AC ANALOG METERS

Permanent Magnet Moving Coil (PMMC) meter movements do not work correctly if directly connected to alternating current, because the direction of the needle movement changes with each half-cycle of the ac. In order to use a dc meter movement such as the d'Arsonval design, the alternating current must be rectified into dc. Half-wave rectifiers or bridge rectifiers are used for this purpose.

1.4.1 Half-wave Rectifier-based Meter

Figure 1.24 shows the circuit of a half-wave rectifier circuit connected to the movement.

A and B are the input terminals of the ac voltmeter. Pure sinusoidal signal with peak value of V_p is applied at the input. Assuming the diode to be ideal, the current through the meter is shown in the figure. The peak value of the current is given by the following equation:

$$I_p = \frac{V_p}{R_s + R_m}$$

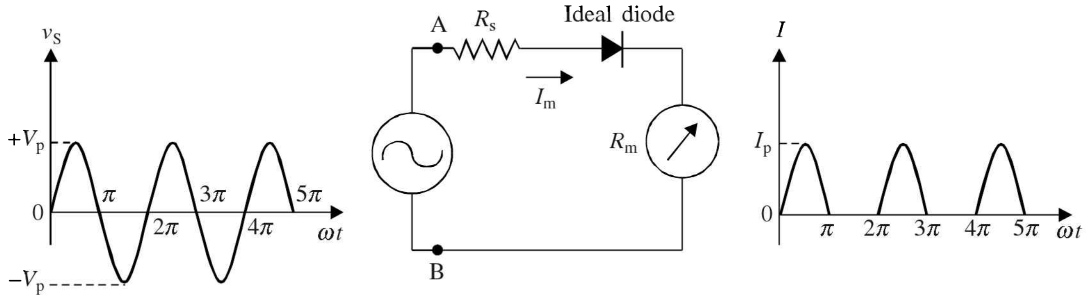


Figure 1.24 Half-wave rectifier-based meter.

If the frequency of the input signal is very low, the deflection of the meter actually follows the variation in current. However, if the frequency is high enough, the inertia of the moving parts prevents the pointer from following the variations in the current. The deflection actually converges upon the average value of the current variation, which is given by the equation,

$$I_{av} = \frac{I_p}{\pi} = 0.318\, I_p$$

Since in ac voltmeters we are interested in measuring the input voltage, the scale is calibrated in volts. Customarily, ac voltmeters are marked in the root mean square (rms) value of the pure sinusoidal input voltage, which is 0.707 of the peak voltage.

To calculate R_s. The multiplier resistance, R_s, can be expressed in terms of the rms voltage, V_{rms}, average current and meter resistance as follows:

$$V_{rms} = 0.707 V_p = 0.707 I_p\,(R_s + R_m) = \frac{0.707}{0.318}\, I_{av}(R_s + R_m)$$

$$= 2.22 I_{av}(R_s + R_m)$$

Therefore,

$$R_s = 0.45 \frac{V_{rms}}{I_{av}} - R_m$$

Note that we have considered an ideal diode. In a practical diode, the diode cut-in voltage and diode resistance should also be considered.

Example 1.10

A 50 μA meter movement with an internal resistance of 50 Ω is to be used in a 0–50 V rms range in the arrangement shown in Figure 1.24. Find the value of the required multiplier resistance.

Solution

$$R_s = 0.45 \frac{V_{rms}}{I_{av}} - R_m = 0.45 \frac{50}{50 \times 10^{-6}} - 50 \cong 450\ \text{k}\Omega$$

The following problems exist in the circuit of Figure 1.24.

(a) In the negative-half cycle of the input, reverse current flows in the circuit, thereby reducing the overall average value of the current. The meter therefore reads low.

(b) In the negative-half cycle, the diode may breakdown due to a large reverse voltage if the same is not of appropriate rating.

The circuit of Figure 1.25 takes care of the above problems. In the negative-half cycle, diode D_2 conducts, thereby shunting all the current. No current therefore flows through the movement. Further, the voltage across any diode when it is reverse biased is much less than the large peak input voltage.

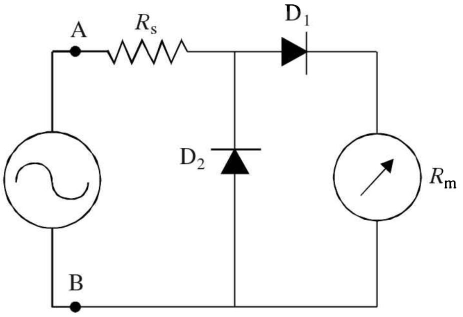


Figure 1.25 A practical half-wave rectifier voltmeter.

In the positive-half cycle, only diode D_1 conducts. This circuit is same as that of Figure 1.24. So, the same analysis is valid.

Sensitivity of the voltmeter

The input resistance of the circuit of Figure 1.25 is $R_s + R_m$ in the positive-half cycle and R_s in the negative-half cycle. Generally, $R_s >> R_m$. We can therefore assume that the internal resistance is R_s throughout the entire cycle. The input resistance can thus be expressed as follows:

$$R_{in} = 0.45 \frac{V_{fsd}}{I_{fsd}} = S_{ac} V_{fsd}$$

I_{fsd} is the full-scale average current and V_{fsd} is the full-scale deflection (rms) voltage. S_{ac} is the ac sensitivity of the voltmeter and is defined as

$$S_{ac} = \frac{0.45}{I_{fsd}}$$

We have seen earlier that dc sensitivity is defined as follows:

$$S = \frac{1}{I_{fsd}}$$

Apparently, ac sensitivity is always less than the dc sensitivity.

Example 1.11

An ac voltmeter has dc sensitivity of 10 kΩ per volt and ac sensitivity of 2 kΩ per volt. Find the input resistance of the voltmeter in the 0–10 V dc range and 0–10 V ac rms range.

Solution

For 0–10 V dc range

$$R_{in} = S_{dc} \times V_{fsd} = 10 \times 10^3 \times 10 = 100\ \text{k}\Omega$$

For 0–10 V ac rms range

$$R_{in} = S_{ac} \times V_{fsd} = 2 \times 10^3 \times 10 = 20\ \text{k}\Omega$$

Since the input resistance for ac is less than that of dc input, we expect a larger loading error for ac measurements than dc measurements.

Nonlinearity on the lower rms ranges

We know that the *I-V* characteristic of a diode is nonlinear. Nonlinearity is maximum at low values of current. To keep the meter scale linear, we should operate the diode in the linear region by operating the diode at relatively high values of current. Figure 1.26 shows an arrangement where the value of R_s is reduced and a resistance, R_{sh}, is connected in parallel with the meter. This increases the current through the diodes.

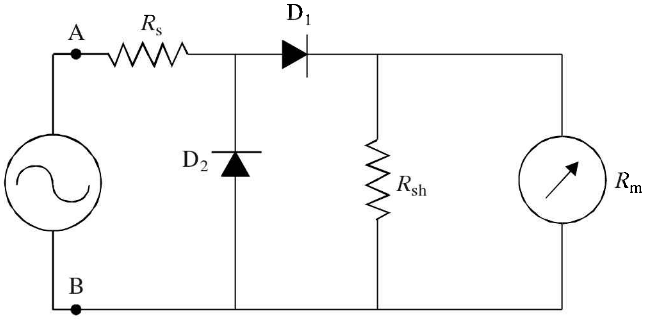


Figure 1.26 Improving the linearity.

If we take $R_{sh} = R_m$, the current through diode D_1 doubles to $2I_{av}$. In general,

$$R_s = 0.45\frac{V_{rms}}{2I_{av}} - (R_D + R_{sh} \parallel R_m)$$

where R_D is the diode forward resistance.

The ac sensitivity is now defined as

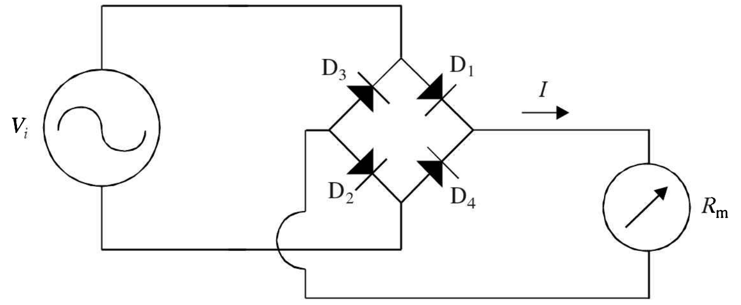
$$S_{ac} = \frac{0.45}{2I_{fsd}}$$

which reduces further.

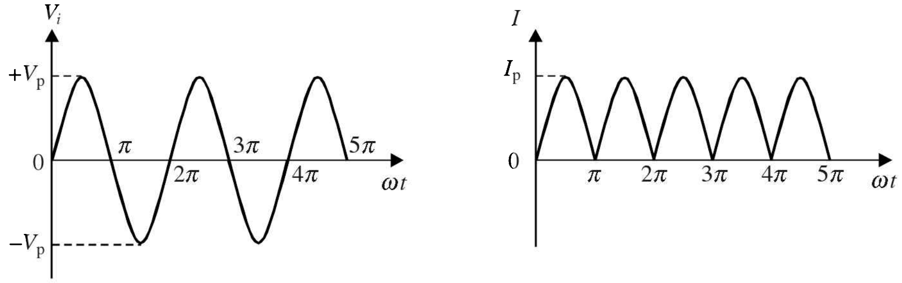
1.4.2 Full-wave Rectifier-based Meter

One way to improve the sensitivity is to use a full-wave rectifier. Figure 1.27(a) shows the circuit with a bridge rectifier.

Diodes D_1 and D_2 conduct for positive-half cycle and D_3 and D_4 conduct for the negative-half cycle of the input. For both the halves, the direction of the current through the meter remains the same as shown in Figure 1.27.



(a) Circuit diagram



(b) Input voltage waveform (c) Meter current waveform

Figure 1.27 Bridge rectifier-based meter.

The average current is given by

$$I_{av} = \frac{2I_p}{\pi} = 0.636I_p$$

The multiplier resistance, R_s, is given by

$$R_s = 0.9\frac{V_{rms}}{I_{av}} - R_m$$

Now, the ac sensitivity becomes

$$S_{ac} = \frac{0.9}{I_{fsd}}$$

which is double that of with the half-wave rectifier.

1.4.3 Nonsinusoidal Waveform Input

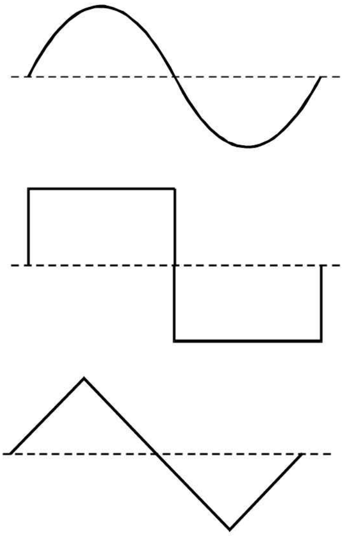
Meter movements relying on the motion of a mechanical pointer tend to mechanically average the instantaneous values into an overall average value for the waveform. This average value is not necessarily the same as the rms value, although many times it is mistaken so. Typically, the assumption is made that the waveform to be measured is a sinusoidal one, and then the meter movement scale is altered by the appropriate multiplication factor. This factor is called Form Factor (FF) and is defined as follows:

$$\text{Form Factor} = \frac{\text{rms value of the waveform}}{\text{average value of the waveform}}$$

For a sinusoidal waveform,

$$\text{Form Factor} = \frac{0.707\,V_p}{(2/\pi)\,V_p} = 1.11$$

Figure 1.28 shows the rms, average and peak-to-peak values of sinusoidal, rectangular and triangular waveforms varying from –1 V to +1 V.



rms = 0.707(peak)
avg = 0.637(peak)
p-p = 2(peak)

rms = peak
avg = peak
p-p = 2(peak)

rms = 0.577(peak)
avg = 0.5(peak)
p-p = 2(peak)

Figure 1.28 The rms, average and peak-to-peak values of sinusoidal, rectangular, and triangular waveforms.

The form factors of rectangular and triangular waveforms are therefore as follows:

$$\text{Form Factor} = \frac{1\,\text{V}}{1\,\text{V}} = 1.00 \quad \text{(rectangular waveform)}$$

$$\text{Form Factor} = \frac{0.577\,\text{V}}{0.5\,\text{V}} = 1.154 \quad \text{(triangular waveform)}$$

The ac voltmeter calibrated for the value of pure sinusoidal waveform (FF = 1.11) input will read low for a triangular waveform input (FF = 1.154) and read high for a rectangular waveform (FF = 1.00) inputs.

1.4.4 Peak Reading Voltmeter

Peak detector is used to construct a peak reading voltmeter. Figure 1.29 shows an unloaded peak detector circuit.

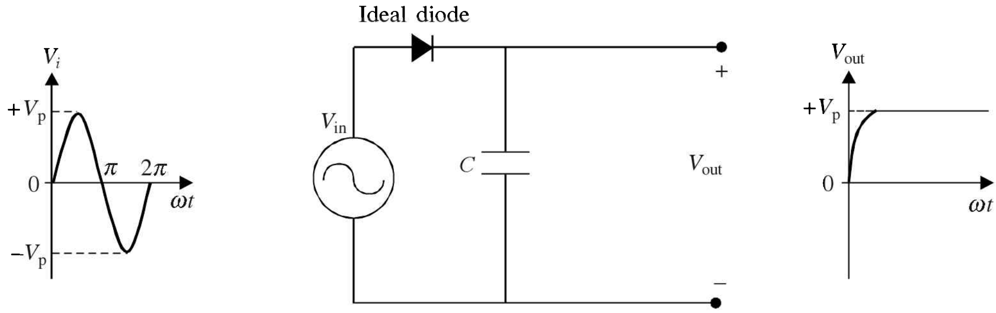


Figure 1.29 Unloaded peak detector circuit.

During the first quarter cycle of the input, diode is forward biased and capacitor charges towards its peak voltage, V_p. At the end of the quarter cycle, the capacitor voltage is V_p. During the next quarter cycle and later, the diode is reverse biased. Hence the output voltage remains constant at V_p. This is shown in the Figure 1.28.

This is not a practical circuit, since on removal of the input signal the capacitor would still remain charged to the peak voltage. If the peak voltage of the input reduces, the capacitor will still indicate the original voltage V_p to which it has been charged.

If we add a resistor, R, of large value parallel to the capacitor (as shown in Figure 1.30), The operation of the circuit remains more or less the same but the capacitor now gets a discharge path through R. The output shown has small ripple but can be assumed to be a dc whose value is equal to the peak value of the input.

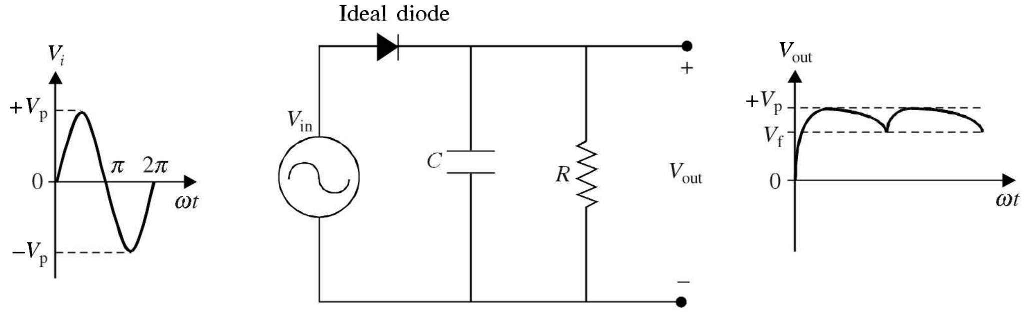


Figure 1.30 Loaded peak detector circuit.

Now if the input voltage reduces, the capacitor discharges through R and attains the new peak voltage.

The output of the peak detector is connected to a dc voltmeter to constitute a simple peak reading voltmeter (see Figure 1.31). The scale is calibrated in terms of the peak voltage.

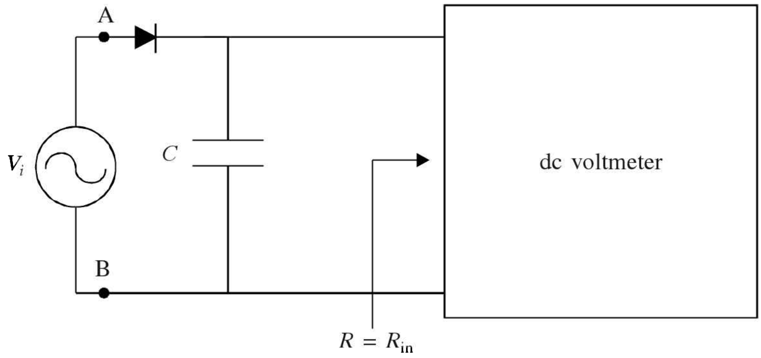


Figure 1.31 Peak reading voltmeter.

Relationship between frequency, R and C

As shown in Figure 1.30, the capacitor discharges from voltage V_p to voltage V_f in approximately one time period, T. The capacitor discharge equation is as follows:

$$V_f = V_p\, e^{-T/RC}$$

We can define accuracy as

$$\frac{V_f}{V_p} = e^{-T/RC}$$

since it relates the final capacitor voltage at the end of the discharge to the peak voltage we intend to measure. Ideally, the right-hand side should be unity. For small values of T/RC, we can write

$$\frac{V_f}{V_p} = e^{-T/RC} \cong 1 - \frac{T}{RC} = \text{accuracy}$$

This is the relationship between accuracy, time period and values of R and C. For example, for an accuracy of 0.99, $T/RC = 0.01$. For an accuracy of 0.95, $T/RC = 0.05$. Therefore, for a 99% accurate peak detector

$$RC = 100T = \frac{100}{f}$$

and for a 95% accurate peak detector

$$RC = 20T = \frac{20}{f}$$

The dc voltmeter actually reads the average value between V_p and V_f. Thus for a peak voltage of 100 V and an accuracy of 95%, the output voltage varies from 100–95 V and the dc voltmeter will read 97.5 V.

Example 1.12

For the circuit of Figure 1.31, the input resistance of the dc voltmeter is 10 MΩ. Find the value of the capacitor so that the capacitor voltage decays no more than 1% from 20 Hz to 20 kHz.

Solution

The lowest frequency is the most critical as far as the capacitor discharge is concerned. Therefore,

$$RC = \frac{100}{f} \quad \text{or} \quad C = \frac{100}{Rf} = \frac{100}{10 \times 10^6 \times 20} = 0.5\ \mu\text{F}$$

Effect of source and diode resistance

In the above discussion we assumed that the diode is ideal and the source resistance is zero. In practical circuits, a diode offers forward and reverse resistances. But compared to other resistances, the forward resistance can be considered negligible and the reverse resistance very large.

The source resistance, R_s, however, may not be too small. This affects the charging time constant and the maximum voltage obtainable across the capacitor. These are:

$$\text{Charging time constant} = (R_s \,\|\, R)C$$

$$V_{\text{max}} = \frac{R}{R + R_s} V_p$$

For a good peak detector, $R >> R_s$. We can thus have the following guidelines:

(a) In order for the capacitor to charge to 99% of V_p, R should be greater than $100R_s$.

(b) In order for the capacitor to charge to 95% of V_p, R should be greater than $20R_s$.

Frequency response

The frequency dependence of PMMC instruments with rectifiers depends on the type of the instrument (ammeters are in general less frequency dependent than voltmeters) and on the measurement range chosen. Frequency dependence of rectifier voltmeters depends mainly on the parasitic capacitances of the series resistors, used for changing voltage ranges. The higher the voltage range, the higher the influence of the parasitic capacitance of the series resistor. Therefore, when the frequency is increased while keeping the input voltage value constant, the current through the parasitic capacitance (connected in parallel to the series resistor) is also increased. Since the resulting current flows through the meter movement in series with the resistor, meter deflection is larger than it should be. For low-voltage ranges the influence of the inductance of the series resistors may prevail and the frequency errors may be negative (meter reads less than it should). For higher frequencies too, the parasitic capacitances of the diodes used in the rectifier or the non-zero recovery times of the diodes used may cause frequency errors of the meter. Rectifier voltmeters and ammeters using germanium or silicon diodes as rectifying elements may have frequency bands of the order of tens of kHz.

1.4.5 Peak-to-Peak Reading Voltmeter

The dc clamping circuits are used as peak-to-peak detectors. A negative dc clamping circuit is shown in Figure 1.32.

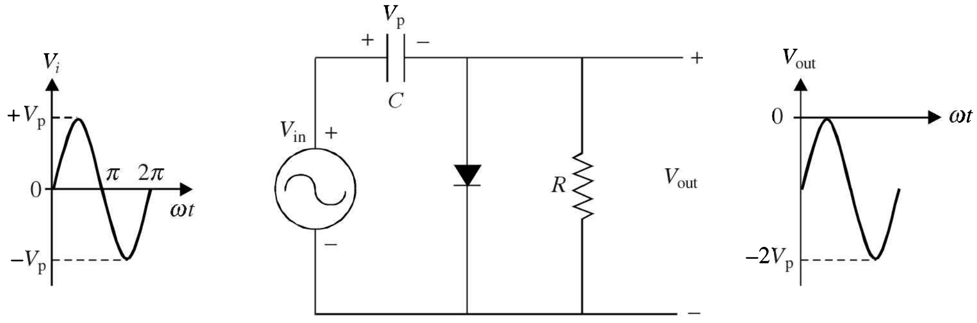


Figure 1.32 Negative dc clamping circuit.

Assuming the diode to be ideal, during the first quarter cycle of the input voltage, the diode is shorted. The capacitor charges towards the peak value of the input V_p and at the end of the quarter cycle reaches V_p. After the first quarter cycle, the capacitor tries to discharge back into the generator. As a result the diode opens and the capacitor discharges through the resistor, R. If R is very large, discharge of the capacitor is minimum and the voltage across the capacitor practically remains V_p. The output is given by the following equation:

$$V_{out} = V_{in} - V_p$$

We get the output which has peak voltage of $-2V_p$. This is called the negative clamper (waveform shifts down) or the negative voltage doubler. Note that the output is independent of the wave shape of the input.

By changing the direction of the diode, we can construct a positive clamper where the output peak will be at $+2V_p$.

A large RC time constant is required at the lowest frequency of operation to operate the circuit properly.

Using the clamper and the peak reading voltmeter, we can construct a peak-to-peak reading voltmeter. Figure 1.33 shows one arrangement with the positive clamper.

Peak-to-peak reading voltmeters are preferred for signals whose negative and positive half cycles are not equal. For example, if there is a square wave with a positive peak of 15 V and a negative peak of 30 V, the peak reading voltmeter will read 15 V, which is misleading. The peak-to-peak reading voltmeter will read 45 V, which is correct.

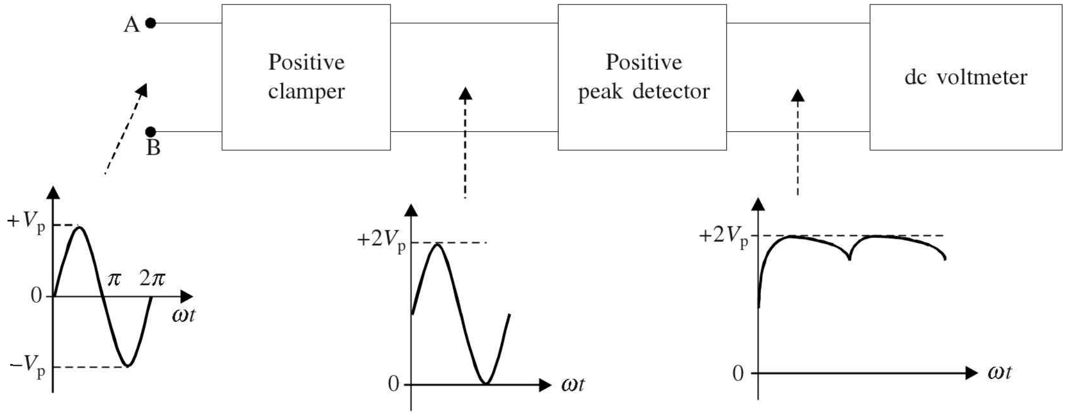


Figure 1.33 Peak-to-peak reading voltmeter.

1.5 OTHER METER MOVEMENTS

The d'Arsonval meter movement (PMMC) is only one type of meter movement. Other types of meter movements can also be used for ac or dc measurements without the use of a rectifier. Two such movements, namely electrodynamic meter movement and moving iron movement are used in some applications. We shall briefly study these movements.

1.5.1 Electrodynamic Meter Movement

An electrodynamic movement uses the same basic operating principle as the basic moving coil meter movement, except that the permanent magnet is replaced by fixed coils (see Figure 1.34). A moving coil, to which the meter pointer is attached, is suspended between two field coils and

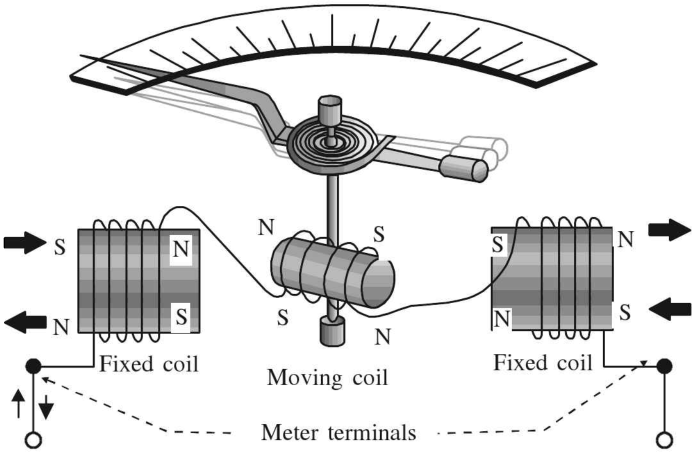


Figure 1.34 Electrodynamic meter movement.

connected in series with these coils. The three coils (two field coils and the moving coil) are connected in series across the meter terminals so that the same current flows through each.

Current flow in either direction through the three coils generates a magnetic field between the field coils. Current flow through the fixed coils sets up a magnetic field in one direction while the current flow through the moving coil sets up an opposing magnetic field. With two magnetic fields opposing each other, the pointer deflects to the right. If the current reverses direction, the magnetic fields of both sets of coils are reversed. With both fields reversed, the coils still oppose each other, and the pointer still deflects to the right. So, this type of meter can be used to measure both ac and dc if the scale is changed. While some voltmeters and ammeters use the electrodynamic principle of operation, the most important application is in the wattmeter.

Deflecting torque

Let I_1 be the current through the fixed coils. Then

$$\text{Flux density, } B = KI_1 \text{ Wb/m}^2 \text{ (tesla)}$$

If I_2 is the current through the moving coil, then

$$\text{Force applied} = NBI_2L \text{ newton}$$

where N is the number of turns in the coil, L is the length of the coil.

And deflecting torque

$$\begin{aligned} T_D &= \text{Force applied} \times \text{perpendicular distance} \\ &= NBI_2Lb \text{ newton-meter} \end{aligned}$$

where b is the breadth of the coil.

For $I_1 = I_2 = I$,

$$T_D = BANI^2$$

where A is the effective coil area ($L \times b$) in m^2.

Since B, A and N are fixed for a given coils, T_D is directly proportional to the square of the current through the coil, i.e. I^2.

Spring controlled torque

$$T_C \propto \text{deflection, } \theta$$

As $T_C = T_D$

$$\theta \propto I^2$$

Deflection is therefore proportional to the square of the current. Therefore, the scale is nonlinear.

Damping

The meter is mechanically damped by means of aluminum vanes that move in enclosed air chambers (see Figure 1.35).

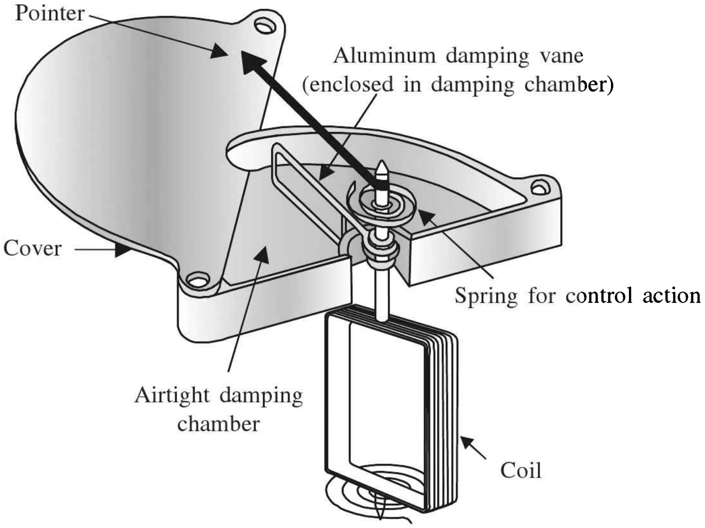


Figure 1.35 Air damping system.

Calibration

For an ac signal, as T_D is proportional to instantaneous current squared which is always positive, torque pulsations are produced. However, the movement because of its inertia, cannot follow these rapid variations of the torque and the pointer takes up the position in which the average torque is balanced by the controlling torque provided by the springs. Therefore, meter deflection is a function of the squared current. The scale is usually calibrated in terms of the rms value of the sinusoidal signal. As the rms value of the dc is its dc value itself, the scale calibrated using dc signal can be used directly to indicate the rms value of the ac signal. This is called the *transfer property*. Hence, the electrodynamometer type of instruments can be calibrated with dc and used to for measuring ac.

Features

Although very accurate, electrodynamometer-type meters do not have the sensitivity of the d'Arsonval-type meter movement.

The primary advantage of the electrodynamometer-type meter movement is that it can be used to measure alternating as well as direct current. No rectifying devices are required to enable the electrodynamometer meter movement to read both ac and dc.

However, the current in the coil has to be limited to prevent excessive heating of the coil which results into increase in resistance. So only a limited current is made to flow through the coil.

The other disadvantages of these meters are, low torque-to-weight ratio resulting in high frictional torque and mechanically fragile construction, making these meters vulnerable to mechanical shocks and vibrations.

Construction of a voltmeter

Similar to the design of a voltmeter with PMMC, here also multiplier resistors are used in series with the coils to limit the current. Multiplier resistors are used depending upon the range of the input voltage.

Construction of an ammeter

Like PMMC movement, this movement also is a basic current reading movement. For high current measurements, shunts are used which are connected in parallel with the moving coil. Shunts are chosen depending upon the range of the input current.

Construction of a wattmeter

A wattmeter is used to measure power. To construct a wattmeter, the fixed (current) coils of the wattmeter are connected in series with the circuit (load), and the voltage coil is connected across the line. When line current flows through the current coil of a wattmeter, a field is set up around the coil. The strength of this field is in phase with and proportional to the line current. The voltage coil of the wattmeter generally has a high-value resistor connected in series with it. The purpose of this connection is to make the voltage coil circuit of the meter as purely resistive as possible. As a result, current in the voltage circuit is practically in phase with the line voltage. Therefore, when voltage is impressed on the voltage circuit, current is proportional to and in phase with the line voltage. The resultant deflection is proportional to the power absorbed by the circuit.

Construction of a watt-hour meter

The watt-hour meter is an instrument for measuring energy, which is the product of power and time. The wattmeter is modified to construct a watt-hour meter for single phase input. It has a small motor (which forms the voltage coil) whose instantaneous speed is proportional to the power passing through it. The motor turns a single rotating disc inside the meter called the *rotor*. This disc in turn drives a series of gears that move dials on a register. These dials are what you look at to determine the amount of electricity consumed. The total revolutions in a given time are proportional to the total energy, or watt-hours, consumed during that time.

1.5.2 Moving Iron (Vane) Meter Movement

The moving iron (vane) meter movement is used for ac meters. It operates on the principle of magnetic repulsion between like poles (see Figure 1.36). The current to be measured flows through a coil, producing a magnetic field, which is proportional to the strength of the current. Suspended in this field are two iron vanes—one in a fixed position and the other, attached to the meter pointer, is movable. The vanes are radical strips of iron. The strips are placed within the coils. The fixed vane is attached to the coil and the movable one to the spindle of the instrument. The magnetic field magnetizes these iron vanes (by eddy currents) with the same polarity regardless of the direction of current flow in the coil. Since like poles

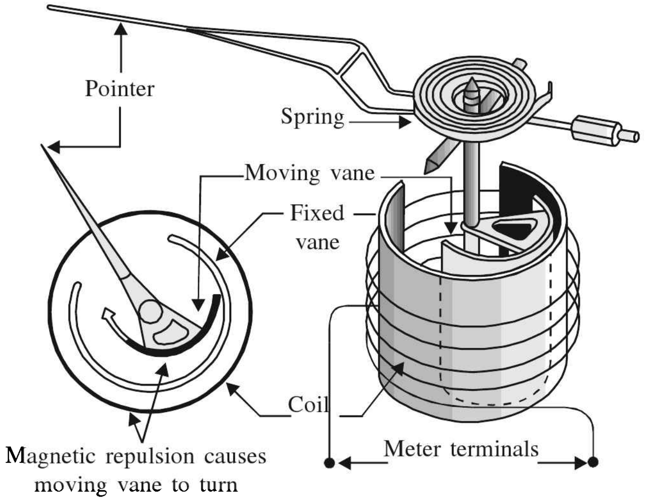


Figure 1.36 Moving iron movement.

repel, the movable vane pulls away from the fixed vane, moving the meter pointer. This motion exerts a torque against the spring. The distance the vane will move against the force of the spring depends on the strength of the magnetic field, which in turn depends on the coil current.

Deflecting torque

Deflecting torque, $T_D \propto$ (pole strength of moving iron, m_1) $\times$ (pole strength of fixed iron, m_2)

$$\propto H \times H \qquad (H \text{ is the magnetic field})$$

$$\propto I \times I = I^2$$

T_D is directly proportional to the square of the current through the coil, i.e. I^2.

Spring controlled torque

$$T_C \propto \text{deflection}, \theta$$

As $T_C = T_D$, we have

$$\theta \propto I^2$$

Deflection is thus proportional to the square of the current. Therefore, the scale is nonlinear.

Damping

The meter is mechanically damped by means of aluminum vanes that move in enclosed air chambers (see Figure 1.35).

Features

These meters are generally used at 50 Hz, but may be used at other ac frequencies as well. By changing the meter scale to indicate dc values rather than ac rms values, moving iron meters can be used to measure dc current and dc voltage.

The accuracy of these meters is limited mainly by the nonlinearity of the magnetizing curve of the iron vanes. At low currents, greater displacement is produced per unit current by peak of the ac than that by the average value, resulting in AC readings higher than the equivalent readings at low end of the scale. At high end of the scale, as the knee of the magnetization curve is produced, the peak value of the ac produces less deflection per unit current than that by the average value, resulting in ac readings lower than the equivalent dc values.

As the reactance of the coil and the magnitude of eddy currents depend upon frequency, the change in frequency may also produce errors.

The meter requires much more power than the d'Arsonval meter to produce a full-scale deflection, thereby reducing the meter's sensitivity.

1.6 ELECTRONIC ANALOG METERS

The voltmeters, ammeters and multimeters discussed above are rugged but suffer from certain disadvantages such as:

- Low sensitivity
- Low input impedance

For example, for a 100 Ω, 50 μA PMMC movement, the sensitivity is 20 kΩ/V. So even for a 0.5 V range, the input resistance is only 10 kΩ. Further, we cannot accurately measure currents less than 50 μA and voltages less than 5 mV (50 μA × 100 Ω). Power absorbed by this movement (0.25 μW) may also be quite large for many electronic circuits. Therefore, an amplifier is necessary for sensitive measurements. The amplifier must have very large input impedance so that it draws minimum current from the electronic circuit. Electronic meters are designed to overcome the problems with the non-electronic meters. Two types of electronic meters are available—electronic analog meters and digital meters. Electronic analog meters are discussed in this chapter and digital meters will be discussed later.

1.6.1 Electronic Analog DC Voltmeter

Figure 1.37 shows the block diagram of an electronic voltmeter.

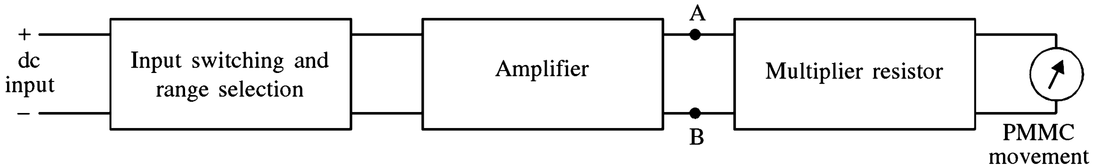
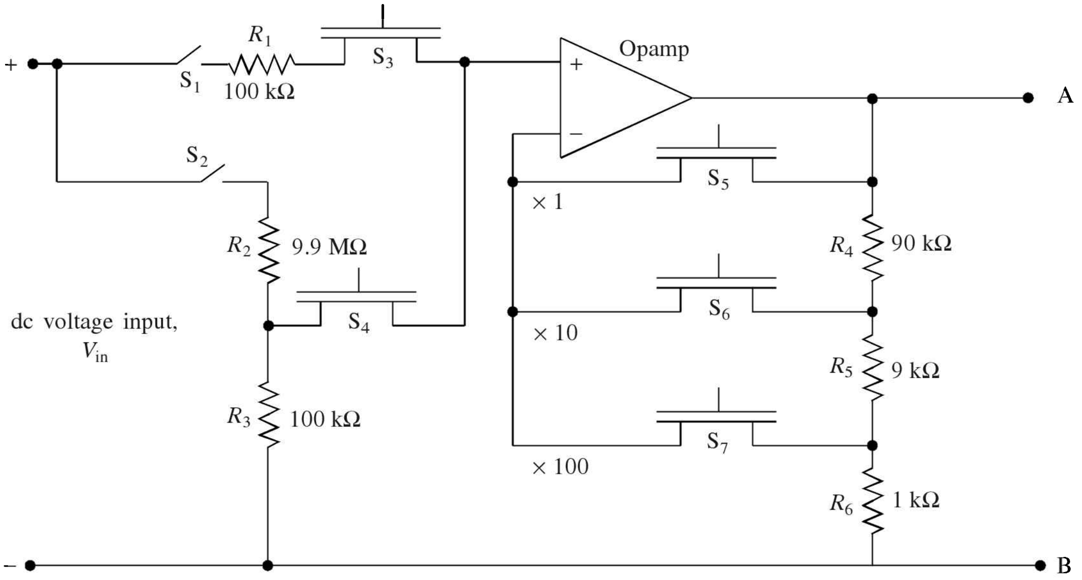


Figure 1.37 Block diagram of an electronic dc voltmeter.

The circuit to the right side of AB is the same as that of a non-electronic dc voltmeter (Figure 1.16). Let us assume that the multiplier resistor is designed for a 1 V input. So the maximum voltage across AB is 1 V. The input switching circuit, range selection switches and amplifier gain is chosen in such a way that the voltage across AB does not exceed 1 V. Figure 1.38 shows one arrangement of switching circuits and amplifier gain selection for the input ranges of 10 mV, 100 mV, 1 V, and 100 V. Switches S_1 and S_2 are mechanical switches and S_3 to S_7 are MOSFET switches. The MOSFET switches can be closed or opened by applying an appropriate voltage at the gate terminal of the MOSFET. Opamp (operational amplifier) is also an FET input amplifier that provides input resistance greater than 10 GΩ.



Figrue 1.38 A typical switching and amplifier gain selection circuit.

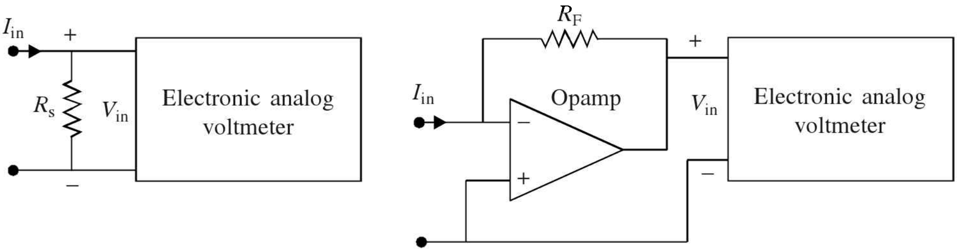
For an input voltage, V_{in}, less than 1 V, switches S_1 and S_3 are closed and the gain of the amplifier is selected so that the voltage across AB does not exceed 1 V. The gain of the amplifier is selected by the resistors R_4, R_5 and R_6. For an input voltage of 10 mV, the gain should be 100, i.e. switch S_7 should be closed (× 100), for an input voltage of 100 mV, the gain should be 10, i.e. switch S_6 should be closed (× 10) and for an input voltage of 1 V, the gain should be 1, i.e. switch S_5 should be closed (× 1).

For an input voltage greater than 1 V, switches S_1 and S_3 are opened and switches S_2 and S_4 are closed. Resistors R_2 and R_3 provide an attenuation of 100. So, the input is divided by 100. Then the gain of the amplifier is selected by closing S_5, S_6 or S_7. For example, a 100 V input is first divided by 100 and then the gain of one (× 1) is selected by closing S_5 to maintain voltage across AB as 1 V.

For the lower input ranges, the input resistance of the voltmeter is essentially same as the input resistance of the amplifier. For the upper voltage ranges, the input resistance is approximately $(R_2 + R_3) = 10$ MΩ.

1.6.2 Electronic Analog DC Ammeter

To construct a dc ammeter, one of the techniques [(a) or (b)] shown in Figure 1.39 is used.



(a) Shunt-type circuit

(b) Circuit for measurement of very low currents

Figure 1.39 Techniques used for construction of ammeters.

In the shunt-type technique of Figure 1.39(a), the input current is converted into voltage by passing the current through a precision resistor, R_s. This voltage, which is proportional to input current, is then measured by the electronic voltmeter. The scale is calibrated in the units of the current. The value of R_s is kept as low as possible to offer minimum input resistance of the ammeter.

For the measurement of very low currents (of the order of pico amperes), the technique of Figure 1.39(b) can be employed where the opamp is being used as a current-to-voltage converter. The input current flows through R_F generating voltage equal to $R_F \times I_{in}$. This voltage is then measured using the electronic voltmeter. The opamp should have low noise and low bias current to measure very low input currents.

1.6.3 Electronic Analog AC Voltmeter

Rectifier-based ac voltmeter

The block diagram of the meter is shown in Figure 1.40. A rectifier is used to convert ac to dc. Since the movement reads average value, the scale is calibrated in terms of rms value of a pure sine wave. To block dc from flowing from the circuit under test to the meter, coupling capacitors are used at various stages.

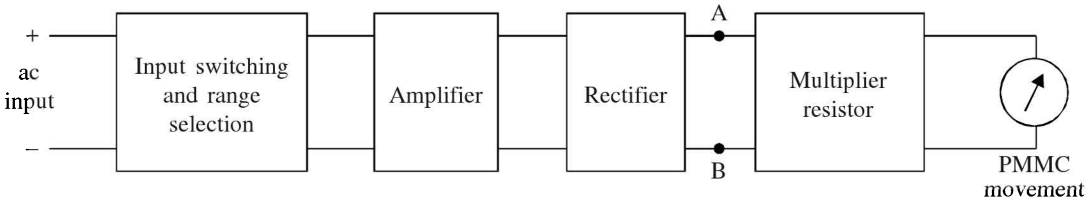


Figure 1.40 Block diagram of an electronic ac voltmeter.

True rms ac voltmeter

The ac voltmeters discussed above are calibrated in terms of rms value of the pure sin wave input. If the input contains harmonics, or the waveform is complex, the measurement with these meters will be wrong.

In order to measure the true rms value of an arbitrary waveform, the concept of power absorption by a heater is used. Power delivered to the heater by a source is given by the following expression:

$$P = \frac{V_{\text{rms}}^2}{R_{\text{heater}}}$$

The heat so generated is measured by a thermocouple (a junction of two dissimilar metals) which generates a voltage proportional to the temperature of the junction. Therefore,

$$V_{\text{th}} = f(P) = f\left(\frac{V_{\text{rms}}^2}{R_{\text{heater}}}\right) = KV_{\text{rms}}^2$$

where K is the constant of proportionality and depends upon the distance between the heater and thermocouple and on the material used in the heater and the thermocouple.

Since the characteristics of a thermocouple are nonlinear, some instruments use another similar thermocouple (balancing thermocouple) for canceling the effect of nonlinearity. Figure 1.41 shows such an arrangement.

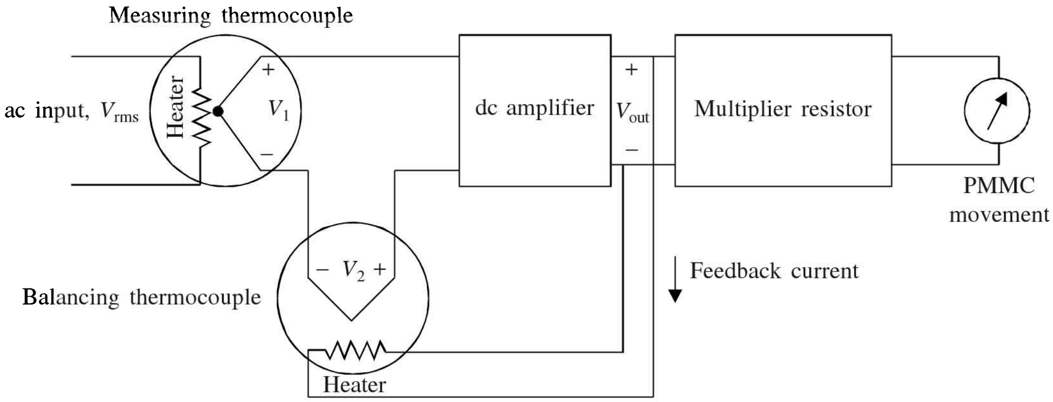


Figure 1.41 Block diagram of a true rms reading voltmeter.

Both the thermocouples are kept in the same environment. The ac input is applied to the heater of the measuring thermocouple, which then generates a dc voltage V_1. Balancing thermocouple voltage V_2 is proportional to the feedback current flowing through the heater. The voltage V_{out} is given by the following equation:

$$V_{\text{out}} = A\,(V_1 - V_2)$$

where A is the voltage gain of the dc amplifier.

Rearranging this equation, we get

$$V_1 - V_2 = \frac{V_{\text{out}}}{A} \cong 0 \text{ for large } A$$

or

$$V_1 \cong V_2$$

or

$$K\,V_{\text{rms}}^2 \cong K\,V_{\text{out}}^2$$

or

$$V_{\text{rms}} \cong V_{\text{out}}$$

Hence, the voltage measured by the voltmeter is approximately equal to rms value of the input signal of any type. If the input voltage is not sufficient to provide the desired heat, the input is first amplified by using an ac amplifier and then fed to the heater. Sensitivities in the millivolt range are possible with such an arrangement.

1.6.4 Electronic Analog AC Ammeter

Shunt type arrangement similar to the dc current measurement is used in these meters. Input current is made to flow through a precision resistor and the voltage so developed across the resistor is measured by an ac voltmeter.

1.6.5 Electronic Analog Ohmmeter

An ohmmeter is designed by using a known constant current source and a voltmeter. Figure 1.42(a) shows the arrangement. The current flowing through the unknown resistor develops voltage across it, which is then read by a voltmeter. The scale is calibrated in ohms. For a multirange ohmmeter, the value of the current source is changed.

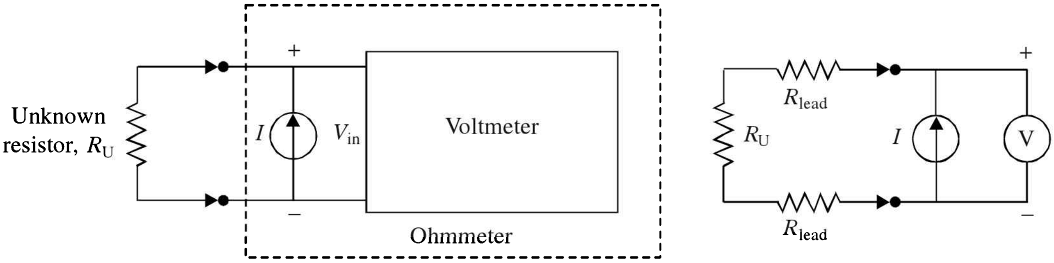


Figure 1.42 (a) Block diagram of an ohmmeter and (b) simplified schematic of two-wire sensing.

Two-wire sensing

The arrangement of Figure 1.42(a) utilizes two-wire sensing. When the same meter terminals are used to measure the voltage drop across the unknown resistor, as are used to supply the current, a meter is said to use a two-wire ohms technique. As shown in Figure 1.42(b), the wire or lead resistances, R_{lead}, which are the resistances due to connecting wires, are added to the unknown resistance value. This causes a large error in the measurement of the unknown resistor, which has the value of the order of lead resistors. The two-wire technique is often used due to its simplicity. The lead resistance errors are removed by using the four-wire ohms technique.

Four-wire sensing

The four-wire sensing resistance technique (also known as the Kelvin sensing technique) provides the most accurate way to measure small resistances. Lead and contact resistances are reduced in this technique. The arrangement is shown in Figure 1.43.

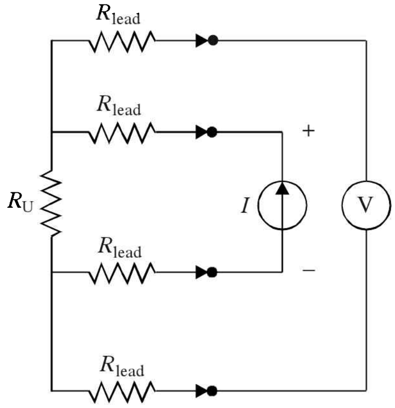


Figure 1.43 Simplified schematic of four-wire sensing.

There are two separate pairs of connections to the unknown resistor. One connection pair, also called "source leads", supplies the current, I, from the constant current source, to the unknown resistor. This current develops voltage across R_U and R_{lead}. The other connection pair, also known as "sense leads", connects the voltmeter with the unknown resistor. The voltmeter is assumed to have very large input resistance. Therefore, no current flows through R_{lead} of sense leads. The voltmeter reads the voltage across R_U, which is equal to $R_U \times I$. This voltage is also independent of R_{lead} of source leads. So this arrangement removes from the measurement the voltage drops in both the source leads and sense leads.

This technique is widely used in applications where lead resistances become quite large and variable. The only disadvantage is in multi-channel systems which require twice as many switches and twice as many wires as the two-wire technique.

1.6.6 Electronic Analog Multimeter

Similar to a non-electronic multimeter, an electronic multimeter also includes the circuits, with a common movement, for the measurement of multirange ac and dc voltages and currents and also the circuits for the measurement of resistances.

Device testing

The resistance measurement circuitry can also be used to test semiconductor devices like diodes and transistors, considering the fact that the semiconductor junction, when forward biased, offers minimum resistance as compared to a reversed biased junction. We have seen that internal battery or the current source is used for the measurement of an unknown resistance. When we connect the positive and negative terminals of the meter with the device electrodes, the junction will either be forward biased or be reverse biased. This resistance value is measured. Then the device electrodes are reversed and again the resistance is measured. Looking at the difference of the two values of these resistances, we can draw a conclusion about the operationality of the device and can even identify the electrodes. Device testing facility is offered by many multimeter manufacturers.

QUESTIONS

1.1 Explain the construction, operation and dynamic behaviour of permanent magnet moving coil (PMMC) mechanism.

1.2 How can the temperature compensation be achieved in a permanent magnet moving coil (PMMC) mechanism?

1.3 What will be the behaviour of a permanent magnet moving coil (PMMC) mechanism when connected across an ac source?

1.4 A basic dc ammeter circuit uses a PMMC meter movement of 1 mA, 100 Ω internal resistance, and an external shunt of 0.001 Ω. What is the range of current the circuit can measure?

1.5 What are the advantages of using the Ayrton shunt?

1.6 How does an ammeter load a circuit? Explain it with an example.

1.7 An ammeter and a voltmeter are connected to a 5 kΩ resistor in a circuit. When the connection is as shown in Figure Q1.7(i), the ammeter and voltmeter readings are 9.98 mA and 50 V respectively. When the connection is as depicted in Figure Q1.7(ii), the respective readings are 10.03 mA and 49.90 V. Find the resistance of the ammeter and the voltmeter. This method can be used to find the unknown resistance and hence the sensitivity of the measuring instruments.

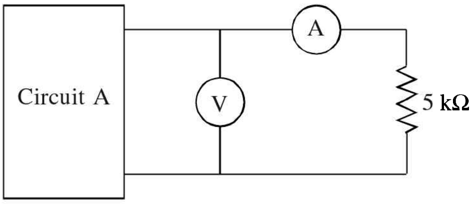


Figure Q1.7(i)

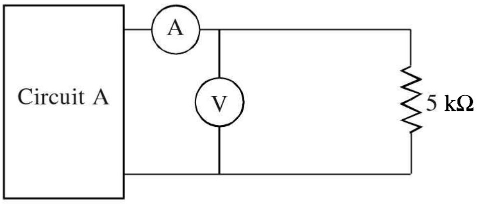


Figure Q1.7(ii)

1.8 In the circuit shown in Figure Q1.8, the voltage across the resistor R_L (= 500 kΩ) is to be measured first by using a voltmeter with sensitivity of 100 kΩ/V and then with a voltmeter of sensitivity of 500 kΩ/V. Calculate the reading of the voltmeter in each case and the % error in the measurement.

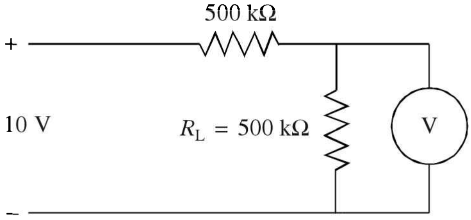


Figure Q1.8

1.9 (a) Figure Q1.9 shows the circuit of a basic series type ohmmeter. Find the ratio of the current I (in the presence of unknown resistance, R_x) to the full-scale deflection current, I_{fsd}.

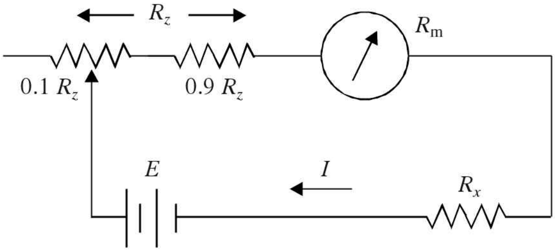


Figure Q1.9

(b) A 1 mA full-scale deflection current PMMC meter movement is to be used in the circuit of Question 1.9(a) with internal resistance, R_m = 100 Ω, and E = 3 V. Calculate the unknown resistance, R_x for 20%, 50% and 100% full-scale current values and mark off the meter face for measurement of resistance.

1.10 In the circuit of Figure Q1.10, a 1 mA meter movement with an internal resistance, R_m = 100 Ω is to be used. The battery voltage is 9 volts. Half-scale deflection should be for R_x = 5000 Ω. (a) Calculate the values of R_1 and R_2. (b) What is the change in the value of R_2 if battery voltage reduces by 10%? (c) What is the half-scale deflection if battery voltage reduces by 10%?

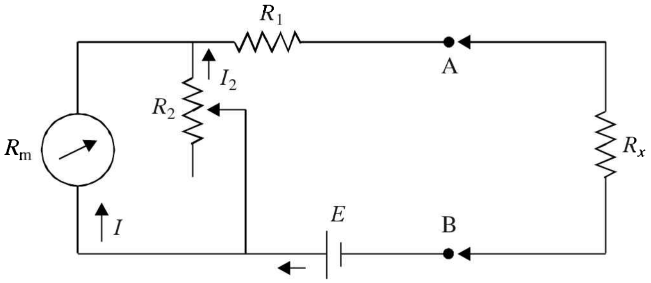


Figure Q1.10

1.11 In the circuit of Figure Q1.11, a 1 mA meter movement with an internal resistance of 100 Ω is to be used. The battery voltage is 3 volts. Half-scale deflection should be for $R_x = 1\ \Omega$. Calculate the values of R_1 and R_{sh}.

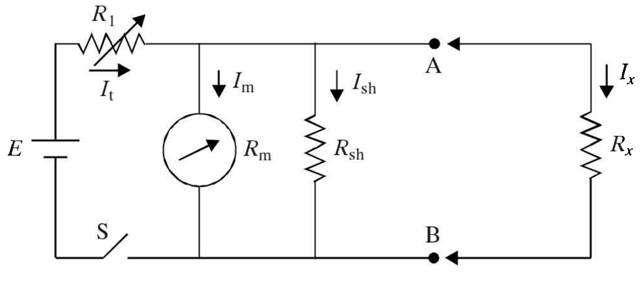


Figure Q1.11

1.12 The ac voltmeter is used to measure voltage V_o across 60 kΩ resistor in the circuit of Figure Q1.12. What is the minimum voltage reading that should be observed? The parameters of the ac voltmeter are:

- Full wave rectification
- 100 μA meter movement
- 15 V range
- Accuracy = ± 3% fsd

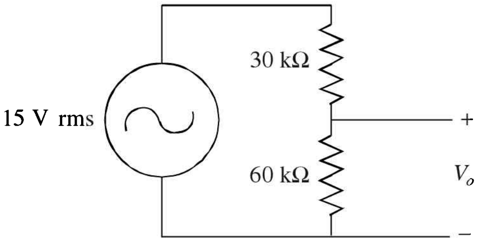


Figure Q1.12

1.13 The voltmeter in Figure Q1.13 uses a 1 mA meter movement with an internal resistance of 100 Ω. The shunt resistance R_{sh}, across the movement is 400 Ω. Diodes D_1 and D_2 each have a forward resistance of 100 Ω, zero cut-in voltage and infinite reverse resistance.

(a) Explain the function of the shunt resistor across the meter movement.

(b) Explain the function of the diode D_2.

(c) Calculate the values of series resistor R_1, R_2, and R_3 if the required meter ranges are 10 V, 50 V, and 100 V, respectively.

(d) Determine the ohm-per-volt rating of this ac voltmeter.

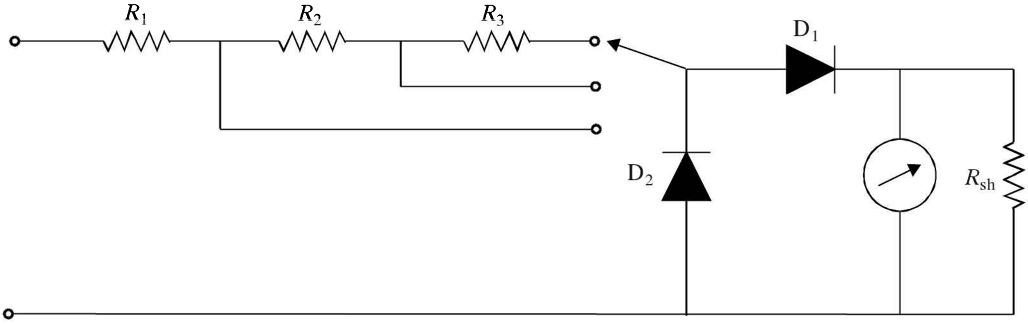


Figure Q1.13

1.14 Explain the construction and operation of a true rms reading voltmeter.

1.15 A current $i = 0.8 + 0.5\sin\omega t - 0.3\sin 2\omega t$ is passed through the circuit of Figure Q1.15. M_1 is a PMMC type current meter with negligible resistance and M_2 is a true rms reading voltmeter with very large resistance. Determine the reading of each meter at $\omega = 10^6$ rad/s.

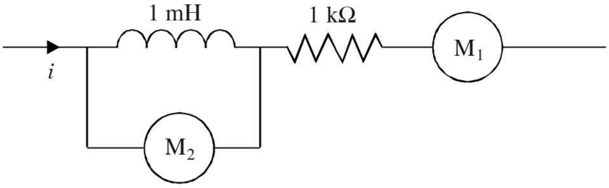


Figure Q1.15

1.16 Figure Q1.16 shows a basic electronic ohmmeter circuit. Assume a meter movement with 500 μA full-scale deflection current and 2 kΩ internal resistance. The unknown resistance is connected across terminals XX′. Calculate the value of V_{DC} and R so that the half-scale deflection reading corresponds to 75 ohms. Consider the opamp as ideal.

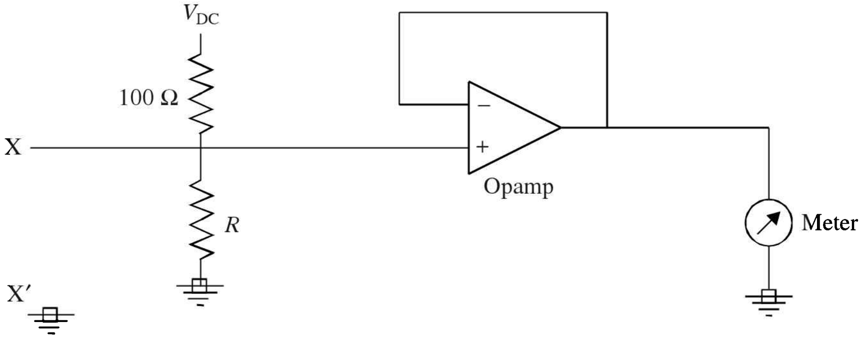


Figure Q1.16

1.17 A sawtooth waveform as shown in Figure Q1.17 is fed to four types of ac electronic voltmeters as follows:

(a) Peak-to-peak reading
(b) Average reading (with full-scale rectification)
(c) Peak reading
(d) True rms reading

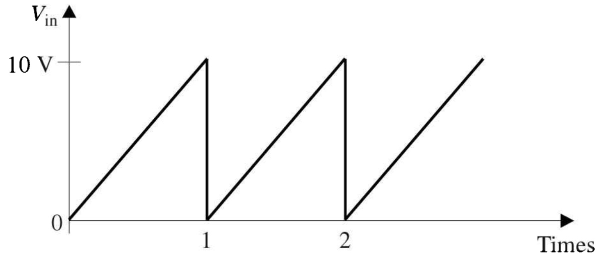


Figure Q1.17

What will be the reading displayed on each voltmeter? What is the percentage error in each case? Note that (b) and (c) types of voltmeters are calibrated for rms values of a pure sinusoidal input.

1.18 A dc voltmeter has been designed with 1 MΩ input resistance and 10 V range. Using this voltmeter, design a 99% accurate peak reading voltmeter between a frequency range of 10 Hz to 10 kHz. What is the maximum peak value of the input, which can be measured by this voltmeter? What factors affect the performance of this voltmeter?

1.19 For the waveform given in Figure Q1.19, will a meter designed to read sinusoids read high or low?

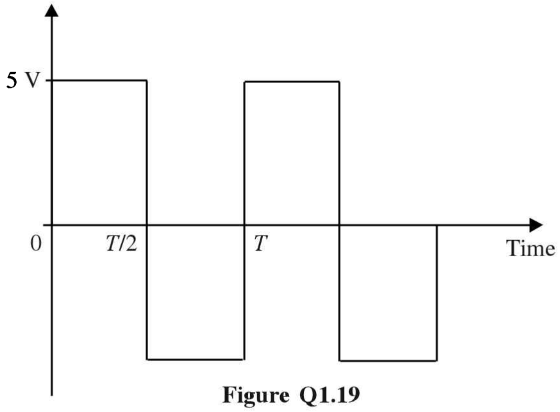


Figure Q1.19

1.20 Design a switching and amplification gain selection circuit for an electronic analog dc voltmeter using FETs and an opamp. The input ranges of the voltmeter should be 10 mV, 100 mV, 1 V, and 10 V. The input resistance of the voltmeter should be 10 MΩ.

1.21 Write typical specifications of an analog multimeter used in the laboratory.

Chapter 2

Digital Meters

2.1 INTRODUCTION

Analog meters use an electromechanical movement of a pointer over a calibrated scale to display the value of the quantity (voltage, current, resistance, etc.) being measured by the meter. Digital meters, on the other hand, use an alphanumeric or numeric LED (Light Emitting Diode) display or LCD (Liquid Crystal Display) for showing the reading of the measured quantity. Further, the internal circuitry used in digital meters is different from that of analog meters. Digital meters normally use analog-to-digital (A/D) converters (or ADC) and other digital processing circuits. Figure 2.1 shows the block diagram of a digital multimeter.

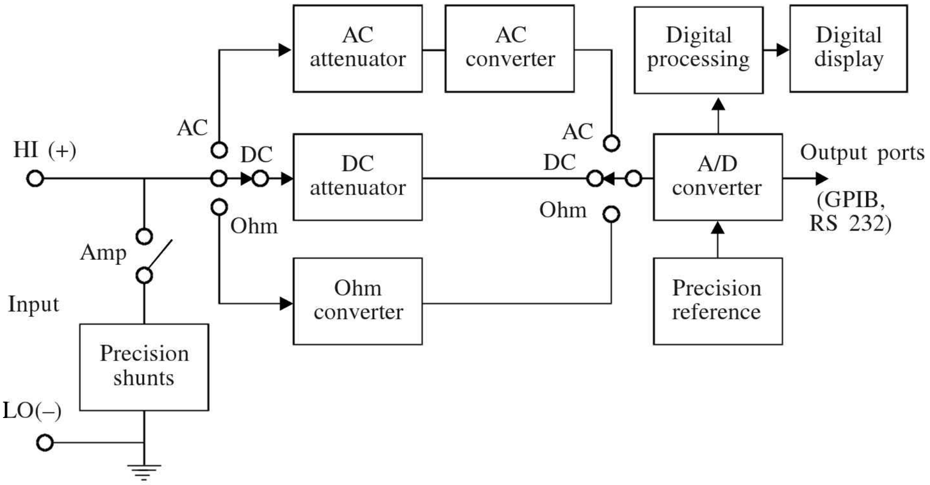


Figure 2.1 Block diagram of a digital multimeter.

The basic input blocks are same as used in an electronic analog multimeter. The attenuators are used for lowering (attenuating) the input voltage. Annexure A gives more details on attenuators. DC input voltage is first attenuated, if required, and then interfaced with an A/D converter which converts it into digital form. This digital signal is then processed and displayed.

AC input voltage is also first attenuated, if required, and then fed to an AC converter which rectifies it. The rectified output is then interfaced with an A/D converter which converts it into digital form. This digital signal is then displayed after processing.

The current is converted to a voltage by passing it through precision shunt resistors. The voltage is then processed and displayed in the units of current.

For ohms measurement, an ohms converter is used. As explained for the electronic analog ohmmeter, the converter includes a constant current source which is made to flow through the unknown resistor. The voltage so developed is then measured and displayed in the units of resistance.

We observe that the A/D converter plays an important role in the digital meters. These converters also constitute the core of other instruments like oscilloscopes, signal analyzers, etc. Many modern meters and other digital instruments use microprocessors for processing digital signals and providing other additional features. In this chapter we will study basics of A/D converters and then describe some commonly used architectures.

Some of the A/D converters use built-in digital-to-analog (D/A) converters for converting digital signals to analog signals. Therefore, first let us learn about the different types of converters.

2.2 DIGITAL-TO-ANALOG CONVERTER

Figure 2.2 illustrates the basic function of a D/A converter.

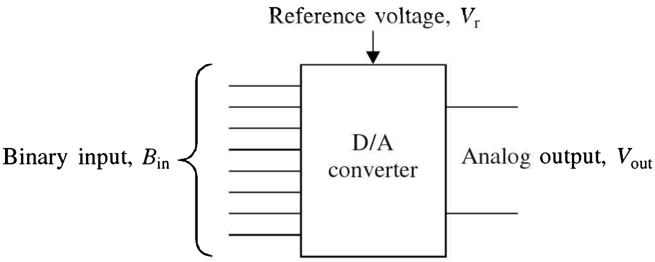


Figure 2.2 Basic function of a D/A converter.

B_{in} is defined to be an N-bit digital word such that

$$B_{in} = b_1 2^{-1} + b_2 2^{-2} + \ldots + b_N 2^{-N}$$

where b_i is a binary digit and equals a 1 or a 0. We also define b_1 to be the most significant bit (MSB) and b_N the least significant bit (LSB). The analog output signal, V_{out}, is related to the digital signal, B_{in}, through an analog voltage reference, V_r. The relationship between these three signals is given by

$$V_{out} = V_r\,(b_1 2^{-1} + b_2 2^{-2} + \ldots + b_N 2^{-N})$$
$$= V_r\,B_{in}$$

We can define V_{LSB} to be the voltage change when one LSB changes, or mathematically,

$$V_{LSB} = \frac{V_r}{2^N}$$

Example 2.1

An 8-bit D/A converter has $V_r = 5$ V. What is the output voltage when $B_{in} = 10110100$? Find also V_{LSB}.

Solution

Using the previous equations, we get

$$B_{in} = 2^{-1} + 2^{-3} + 2^{-4} + 2^{-6} = 0.703125$$
$$V_{out} = V_r B_{in} = 5 \times 0.703125 = 3.516 \text{ V}$$
$$V_{LSB} = \frac{5}{2^8} = \frac{5}{256} = 19.5 \text{ mV}$$

We see that precision depends upon the number of bits N.

2.2.1 Transfer Characteristics

Figure 2.3 shows the transfer characteristics of an ideal 3-bit D/A converter.

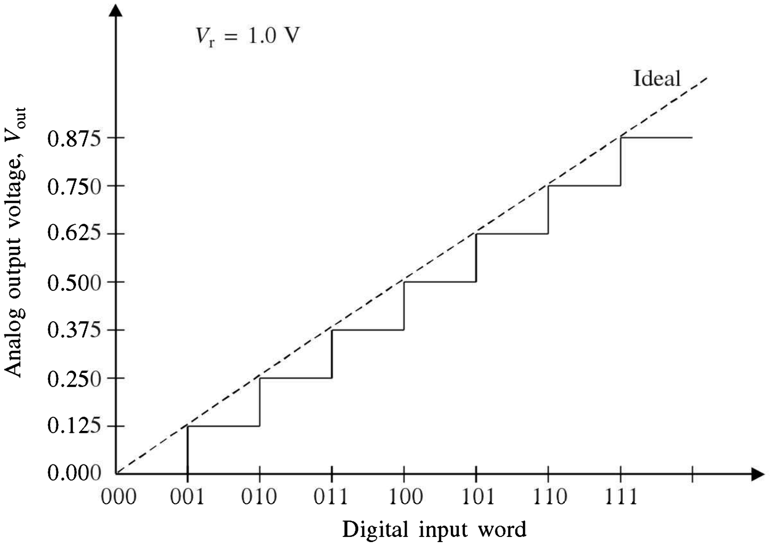


Figure 2.3 Transfer characteristics of an ideal 3-bit D/A converter.

Note that the maximum value of V_{out} is not V_r but rather $V_r(1 - 2^{-N})$, or equivalently $V_r - V_{LSB}$.

In practical circuits, the transfer characteristics are not linear and not monotonic. These are explained in the section on performance parameters of D/A converters.

2.2.2 D/A Conversion Techniques

Various architectures are used to realize a D/A converter. The most commonly used techniques are discussed in the following sections.

Weighted-resistor D/A converter

Figure 2.4 illustrates a 4-bit weighted resistor D/A converter which has a reference voltage source, a set of binary-weighted resistors, an opamp and a set of switches. Each switch has its own binary bit of the digital input word which controls it. The switch is closed when the binary bit is a 1 and open when the binary bit is a 0. The switches are realized using transistors.

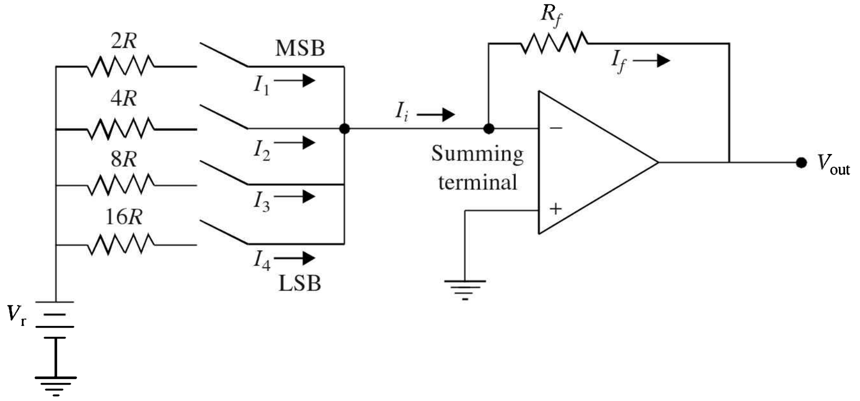


Figure 2.4 A 4-bit binary weighted-resistor D/A converter.

When the binary bit (b) is a 1, the current flows through the resistor in series into the summing terminal. When the currents from the network of resistors are summed in the amplifier, the total current, I_i, is given by

$$I_i = I_1 + I_2 + I_3 + I_4 = I_f$$

The same current flows through the feedback resistor R_f. The output voltage is given by the following equation:

$$V_{out} = -R_f I_f = -R_f V_r \left(\frac{b_1}{2R} + \frac{b_2}{4R} + \frac{b_3}{8R} + \frac{b_4}{16R} \right)$$

$$= -\frac{R_f}{R} V_r (b_1 2^{-1} + b_2 2^{-2} + \ldots + b_N 2^{-N})$$

We see that the output voltage is proportional to the sum of the binary weights. Further, the values of the input resistors are inversely proportional to the input binary weights. The lowest-value resistor (R) corresponds to the MSB and the highest-value ($8R$) to the LSB.

The disadvantage of the weighted-resistor D/A converter is the large number of resistor values that it has. For instance, if we take an 8-bit converter, the 8 resistors will range from R to $128R$ in binary weighted steps. Also, this type of D/A converter is very difficult to mass produce due to the range of resistors required, where the tolerance is less than 0.5%, to accurately convert the input. Further, the lowest value resistor R affects the MSB and must have the highest precision. Also, it draws maximum current from the reference voltage source.

R-2R ladder D/A converter

Unlike the weighted-resistor D/A converter, the R-$2R$ ladder D/A converter uses only two resistor values. Figure 2.5 shows the basic circuit of 4-bit R-$2R$ ladder D/A converter.

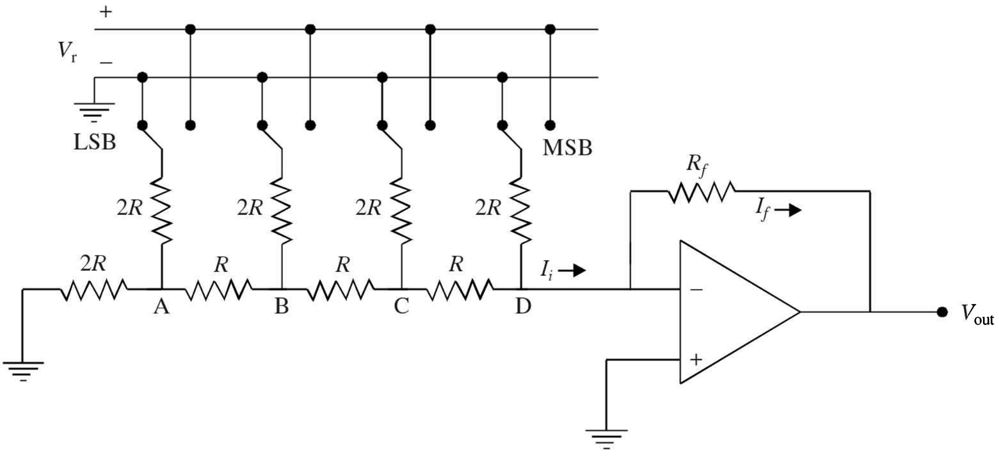


Figure 2.5 A 4-bit R-$2R$ ladder D/A converter.

Each switch has its own binary bit of the digital input word which controls it. The switch is connected to V_r when the binary bit is a 1 and connected to ground when the binary bit is a 0.

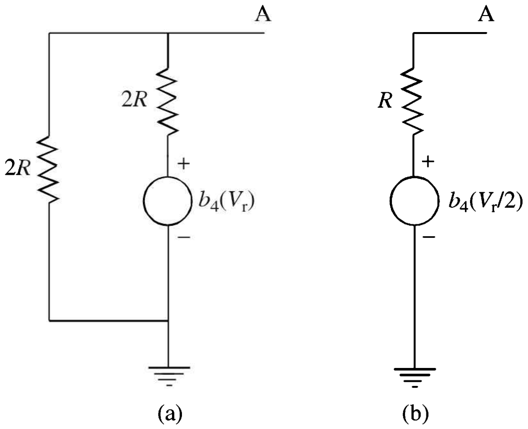
To understand how the circuit operates, consider the first bit (LSB) of the R-$2R$ network. Depending on the state of the switch, this bit forms a voltage divider developing $0 \times V_r$ or $1/2 \times V_r$. In either case, the Thevenin resistance is the same, i.e. R. The Thevenin equivalent to the left of point A is shown in the Figure 2.6(b). Similarly, if we find the Thevenin equivalent to the left of point B, we find that the Thevenin equivalent is as shown in Figure 2.7(b).

We therefore see that the Thevenin equivalent of the circuit to the left of point D would be as shown in Figure 2.8.

From the circuit of Figure 2.5, the current I_i is given by

$$I_i = \frac{[b_1 V_r + (b_2 V_r)/2 + (b_3 V_r)/4 + (b_4 V_r)/8]/2}{R}$$

$$= I_f$$



(a) (b)

Figure 2.6 Thevenin equivalent to the left of point A of Figure 2.5.

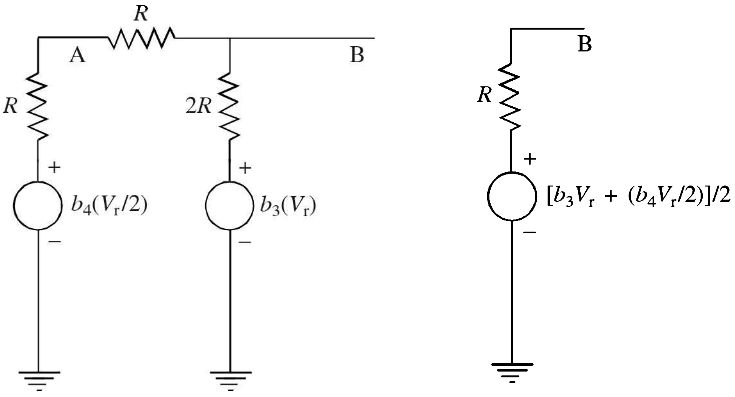


Figure 2.7 Thevenin equivalent to the left of point B of Figure 2.5.

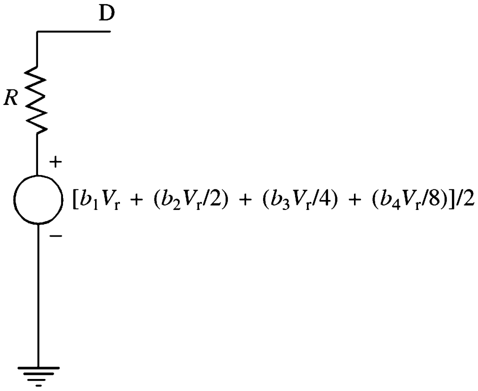


Figure 2.8 Thevenin equivalent to the left of point D of Figure 2.5.

Therefore, the output voltage is

$$V_{out} = -R_f I_f = -R_f V_r \left(\frac{b_1}{2R} + \frac{b_2}{4R} + \frac{b_3}{8R} + \frac{b_4}{16R} \right)$$

$$= -\frac{R_f}{R} V_r \,(b_1 2^{-1} + b_2 2^{-2} + \ldots + b_N 2^{-N})$$

Thus, the R-$2R$ ladder can be used to obtain binary weighted voltages or currents while using only a single-sized resistor. (The resistors of size $2R$ are made of two resistors of size R, to improve matching properties.) As a result, this R-$2R$ approach usually gives both smaller size and better accuracy than the binary-weighted resistor approach. Further, the resistors can be lower in value giving high speed operation.

Inverted R-2R ladder D/A converter

In this configuration, the positions of R and $2R$ resistors are interchanged. The ladder currents are independent of the switch position. This is a popular configuration for CMOS technology.

Switched-capacitor D/A converter

Instead of using resistors in binary weighted-resistor and R-$2R$ ladder networks, capacitors are used, i.e. R is replaced by C. Therefore these networks become binary weighted capacitor and C-$2C$ ladder networks. These converters have low-power dissipation with CMOS technology. CMOS opamps and capacitors are fabricated on the same chip. However, these converters suffer from more dynamic switching losses.

Internal structure of commercial D/A converters

In addition to the resistor network, commercial D/A converters include additional digital and analog circuitry for temporarily storing the digital input and for shifting the voltage levels to get the desired output voltage (see Figure 2.9).

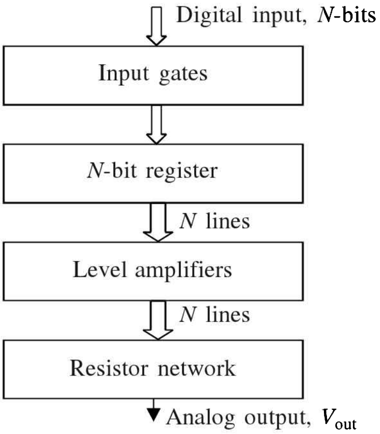


Figure 2.9 Block diagram of commercial D/A converters.

2.2.3 Digital Input Modes of Operation

There are different modes (configurations) that can be configured to operate a D/A converter, e.g. Unipolar Straight Binary (USB), Bipolar Offset Binary (BOB), or Binary Two's Complement (BTC). Table 2.1 shows the meaning of each mode of operation. Depending on the design specifications, one of these three modes may be more useful than the others.

Table 2.1 Modes of operation

Digital input	*Analog output*		
MSB LSB	Unipolar straight binary (USB)	Bipolar offset binary (BOB)	Binary two's complement (BTC)
111111111111	+ Full scale	+ Full scale	– 1 LSB
1000000000000	+ 1/2 Full scale	Zero	– Full scale
011111111111	(+ 1/2 Full scale – LSB)	– 1 LSB	+ Full scale
000000000000	Zero	– Full scale	Zero

2.2.4 Performance Characteristics of D/A Converters

Resolution

It refers to the smallest subdivision (step-size) of the full-scale output that is possible, given the number of bits in the input word. This is equivalent to the weight of one LSB which is defined as

$$V_{\mathrm{LSB}} = \frac{V_r}{2^N}$$

where V_r is the reference (full scale voltage) and N is the number of bits in the input word. Resolution can also be defined as the number of bits that are converted.

Accuracy

Accuracy tells us how close is the output to the ideal converted output. That is, it gives the error in the analog output from the theoretical value for a given digital input. It is generally specified as a percentage of the full scale output. Ideally, the accuracy should be no worse than ± (1/2) LSB. It also includes the *offset* and *gain* errors (see Figure 2.10).

Offset error

The offset error is the ΔV from 0 V that results when a digital code is entered that is supposed to produce 0 V. Related to the offset error is the offset error temperature coefficient, which is the change in offset over temperature. This is usually specified in ppm/°C. Offset error is critical in dc applications. For this reason, a buffer op-amp that does not contribute to the problem must be selected. The opamp's own offset voltage should be much less than that of the converter. In ac applications, the offset error is not important and can be ignored.

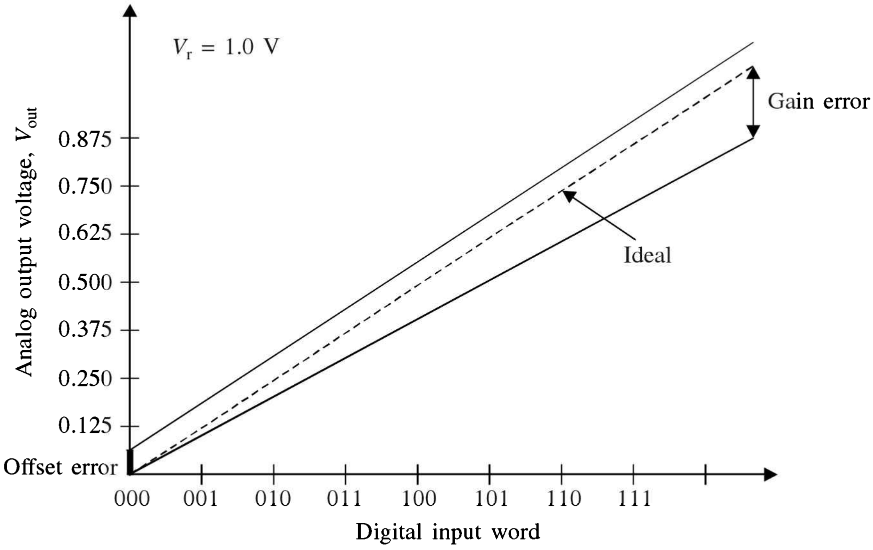


Figure 2.10 Gain and offset errors.

Gain error

The gain of the D/A converter may be greater than or less than the gain needed to produce the desired full-scale analog voltage. The gain error is the difference in slope between the ideal D/A output gain and the actual gain. Related to the gain error is the gain error temperature coefficient, which is the change in gain over temperature. The gain error can be critical in both ac and dc applications.

Example 2.2

Given a 3-bit DAC with a 1 V full scale voltage and accuracy ± 0.2%, find its resolution and accuracy in terms of voltage.

Solution

$$\text{Resolution} = \frac{1}{2^3} = 0.125 \text{ V}$$

$$\text{Accuracy} = (\pm 0.2\%)\ (1 \text{ V}) = \pm 2 \text{ mV}$$

These parameters can be graphically represented as in Figure 2.11.

The offset and gain errors depicted in Figure 2.10 still keep the transfer characteristics linear. In practice, the characteristics, however, may not be linear and may not always increase (be monotonic) with the increase in the input (see Figure 2.12).

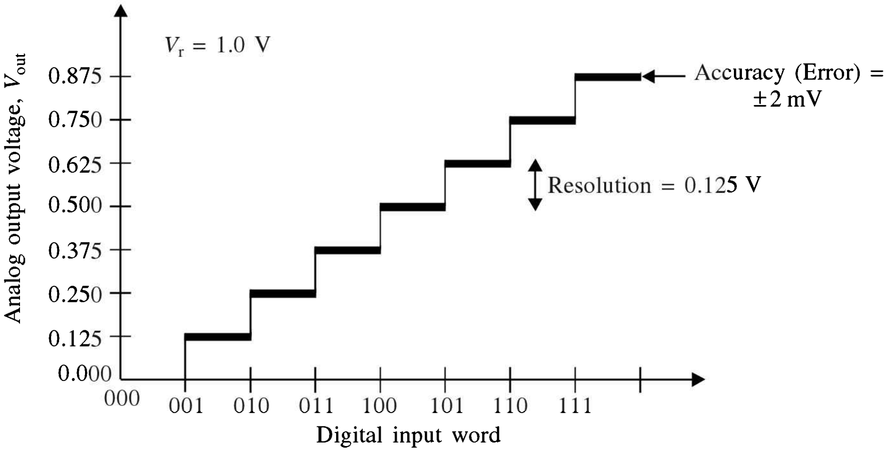
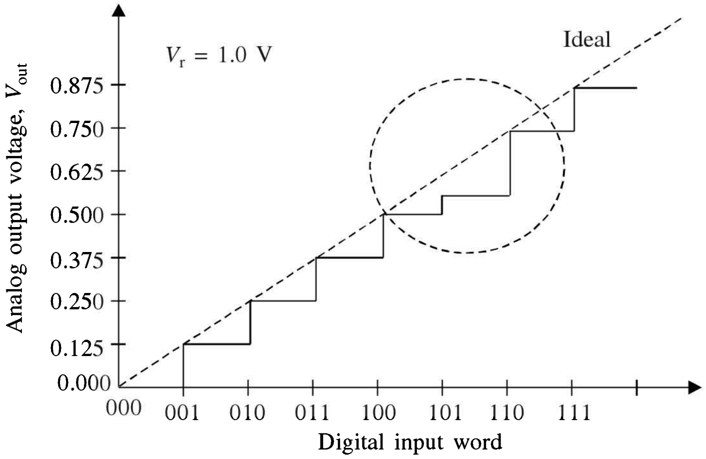
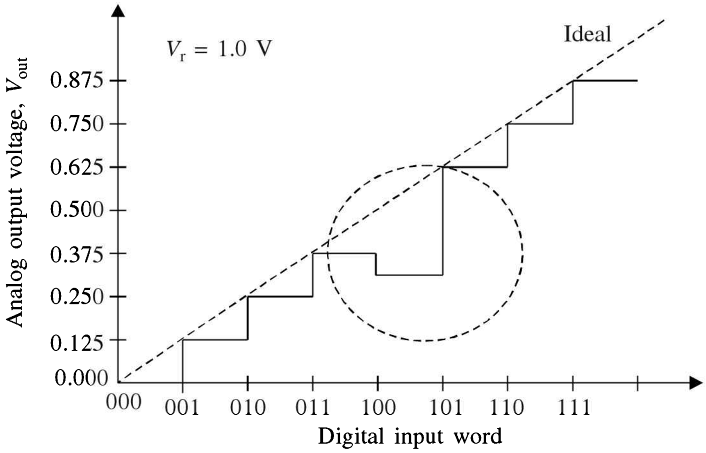


Figure 2.11 Graphical representation of accuracy and resolution.



(a)



(b)

Figure 2.12 Transfer characteristics: (a) Nonlinearity and (b) non-monotonocity.

Besides nonlinearity and non-monotonocity, the following types of error values are also given by the manufacturers:

Differential nonlinearity (DNL) error

The differential nonlinearity error shown in Figure 2.13 (sometimes seen as simply differential linearity) is the difference between an actual step height and the ideal value of 1 LSB. Therefore if the step height is exactly 1 LSB, then the differential nonlinearity error is zero. If the DNL exceeds 1 LSB, there is a possibility that the converter would become non-monotonic. This means that the magnitude of the output gets smaller for an increase in the magnitude of the input. We can define DNL error as follows:

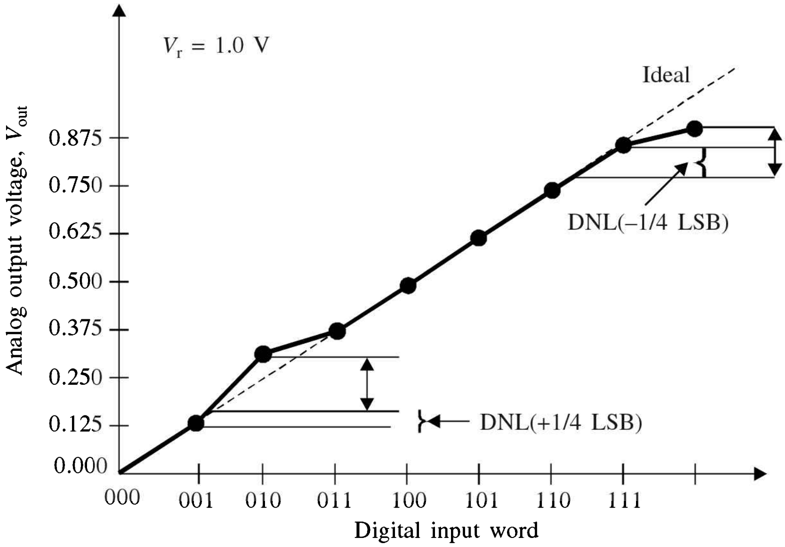


Figure 2.13 Differential nonlinearity error.

$$\text{DNL error} = |V_{out}(i+1) - V_{out}(i)| - V_{LSB}$$

Integral nonlinearity (INL) error

The integral nonlinearity error shown in Figure 2.14 (sometimes seen as simply linearity error) is the deviation of the values on the actual transfer function from a straight line. This straight line can be either a best straight line which is drawn so as to minimize these deviations or it can be a line drawn between the end points of the transfer function once the gain and offset errors have been nullified. The second method is called *end-point linearity* and has the usual definition adopted since it can be verified more directly.

The deviations are measured at each step. The name integral nonlinearity derives from the fact that the summation of the differential nonlinearities from the bottom up to a particular step, determines the value of the integral nonlinearity at that step. Thus,

$$\text{INL error} = \Sigma\,[V_{out}(i+1) - V_{out}(i)] \qquad \text{for } i = 0 \text{ up to the given input}$$

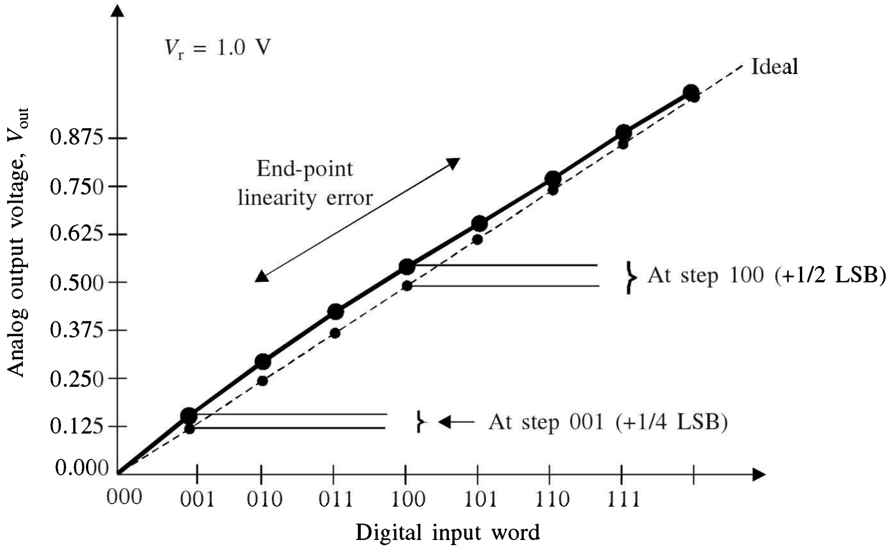


Figure 2.14 Integral nonlinearity error.

Both the INL and DNL errors affect ac applications in the forms of distortion and spectral harmonics (spurs). In dc applications, they result in an error in the dc output voltage.

Settling time

This is the time it takes a D/A converter to settle within an error band around its final value when a change occurs in the digital input.

Dynamic range

This is defined as the number of usable values. For example, a 4-bit D/A converter has a dynamic range of 16 values or $20 \log_{10}16 \cong 24$ dB. Sometimes it is expressed as 6 × number of bits of digital input.

Power supply rejection ratio

The power supply rejection ratio is sometimes called the *power supply sensitivity*. It is the ability of the converter to reject ripple and noise present on its power inputs. DC applications may not be adversely affected. Poor power supply rejection can cause spurs and harmonic distortion in ac applications, as external frequency components leak into the output and modulate with it.

The power supply rejection ratio is defined as the ratio of a per cent of full-scale reading change in the output to a per cent of change in either positive, negative or digital supply voltage about the nominal power supply voltages.

Drift

Many converter parameters like gain and various types of errors change with temperature. This phenomenon is called *drift* and is specified as ppm of FSR/°C with reference to 25°C temperature.

2.3 ANALOG-TO-DIGITAL CONVERTER

Figure 2.15 represents the basic function of an A/D converter.

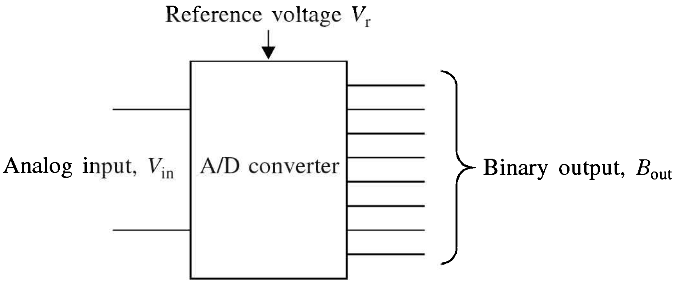


Figure 2.15 Basic function of an A/D converter.

The binary output, B_{out} is the N-bit digital output word while V_{in} and V_r are the analog input and reference signals, respectively. For an A/D converter the following equation relates these signals:

$$V_r\,(b_1 2^{-1} + b_2 2^{-2} + \ldots + b_N 2^{-N}) = V_{in} \pm V_x$$

where b_i is a binary digit and equals a 1 or a 0. We also define b_1 to be the most significant bit (MSB) and b_N the least significant bit (LSB). Also, V_x is given by the following expression

$$-\left(\frac{1}{2}\right) V_{LSB} \le V_x < \left(\frac{1}{2}\right) V_{LSB}$$

where $V_{LSB} = V_r/2^N$.

2.3.1 Transfer Characteristics

The process of converting analog to digital signal is called *quantization*. Note that there is now a range of input values that produce the same digital output word. Some information is lost and some distortion is introduced into the signal during the conversion process. This signal ambiguity produces what is known as *quantization error/quantization noise*. See Figure 2.16. In the case of a D/A converter this effect is not present since the output signals are well-defined.

For the ideal staircase transfer function of an A/D converter [Figure 2.16(a)], the error between the actual input and its digital form has a uniform probability density function if the input signal is assumed to be random. It can vary in the range ±1/2 LSB or $\pm q/2$ where q is the width of one step.

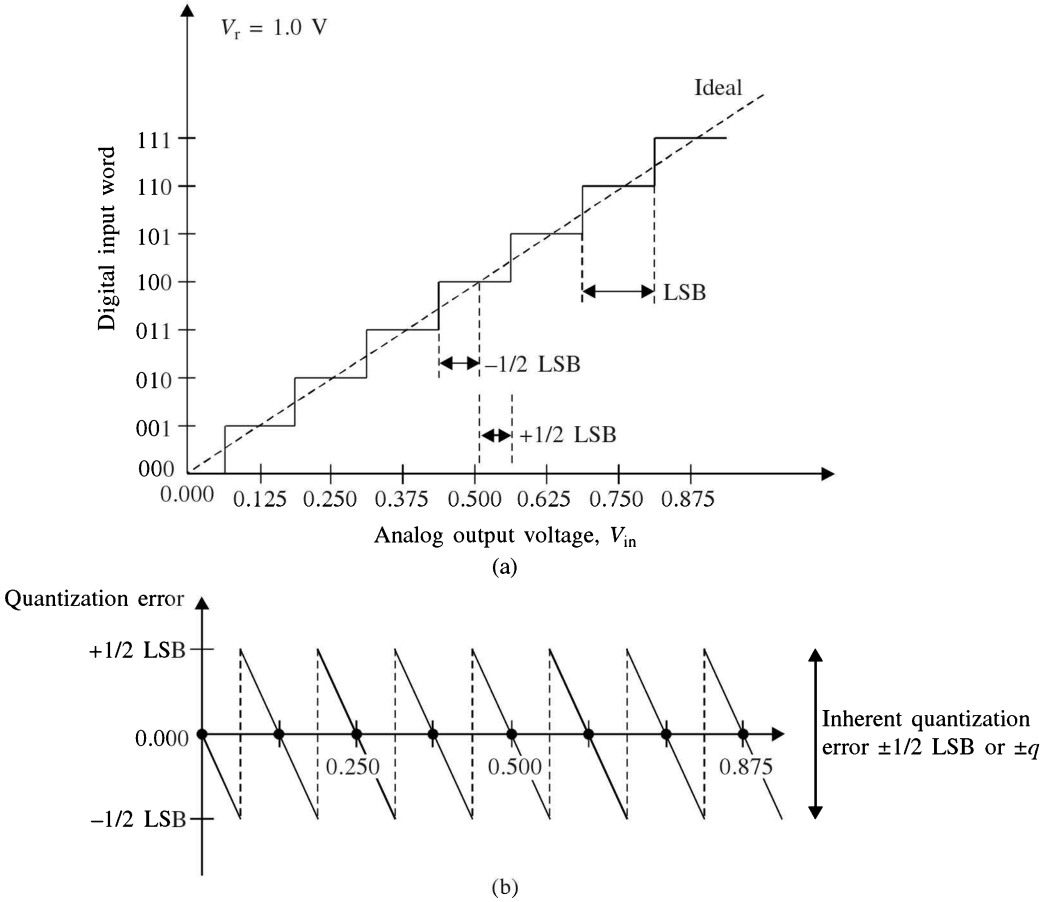


Figure 2.16 Transfer characteristics of an ideal 3-bit A/D converter.

The quantized signal $y[n]$ can be represented by a linear function $Gx[n]$ with an error $e[n]$, that is,

$$y[n] = Gx[n] + e[n]$$

The gain G is the slope of the straight line passing through the center of the quantization characteristic.

To simplify the analysis of noise from quantizer, the following assumptions about the noise process and its statistics are traditionally made:

1. The error sequence $e[n]$ is a sample sequence of a stationary random process.
2. The error sequence is uncorrelated with the sequence $x[n]$
3. The random variables of the error process are uncorrelated, i.e. the error is a white-noise process.
4. The probability distribution of the error process is uniform over the range of quantization error.

For a zero mean $e[n]$, its variance σ_e^2 or e_{rms}^2 (mean square quantization error) is

$$e_{\text{rms}}^2 = \frac{1}{q}\int_{-q/2}^{q/2} q^2 dq = \frac{q^2}{12}$$

When a quantized signal is sampled at frequency f_s, all of its power folds into the frequency band $0 < f < f_s$. Then, if the quantization noise is white, the spectral density of the sampled noise is given by

$$E(f) = \sqrt{\frac{e_{\text{rms}}^2}{f_s/2}} = e_{\text{rms}}\sqrt{\frac{2}{f_s}}$$

The noise power is flat and the level of the noise power spectral density is given by the equation

$$N(f) = \frac{q^2}{12}\frac{1}{f_s}$$

Consider a sine wave input $F(t)$ of amplitude A so that

$$F(t) = A \sin \omega t$$

which has a mean square value of $F^2(t)$, where

$$F^2(t) = \frac{1}{2\pi}\int_0^{2\pi} A^2 \sin^2 \omega t\, dt$$

and which is the signal power. Therefore, the signal to noise ratio, SNR, is given by

$$\text{SNR (dB)} = 10 \log\left(\frac{A^2/2}{q^2/12}\right)$$

But

$$q = 1 \text{ LSB} = \frac{2A}{2^n} = \frac{A}{2^{n-1}}$$

Substituting for q gives

$$\text{SNR (dB)} = 10 \log\left(\frac{A^2/2}{A^2/3 \times 2^{2n}}\right) = 10 \log\left(\frac{3 \times 2^{2n}}{2}\right)$$

$$= (6.02n + 1.76)\text{ dB}$$

This gives the ideal value for an n-bit converter and shows that each extra 1 bit of resolution provides approximately 6 dB improvement in the SNR.

In practice, the errors introduce nonlinearities that lead to a reduction of this value. The limit of a 1/2 LSB differential linearity error is a missing code condition which is equivalent to a reduction of 1 bit of resolution and hence a reduction of 6 dB in the SNR. This then gives a worst case value of SNR for an n-bit converter with 1/2 LSB linearity error as

$$\text{SNR (worst case)} = 6.02n + 1.76 - 6 = (6.02n - 4.24) \text{ dB}$$

This establishes the boundary conditions for the choice of the resolution of the converter based upon a desired level of SNR.

2.3.2 A/D Conversion Techniques

Many architectures are available to realize an A/D converter. Some of these techniques are discussed below.

Single slope (ramp) A/D converter

Figure 2.17 shows the block diagram of this type of converter. It uses a comparison method in which the input is compared with the voltage generated by the ramp generator. The ramp generator is realized using opamp based integrator (see Figure 2.19). The counter counts the pulses and the latches store temporarily the output bits of the counter.

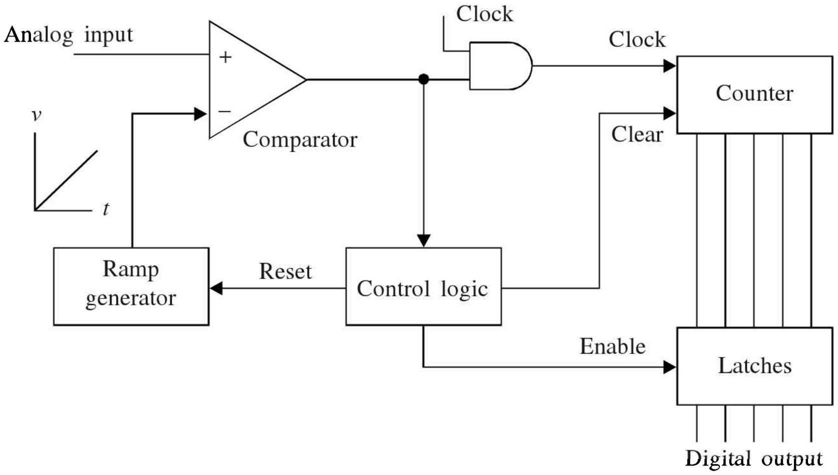


Figure 2.17 Block diagram of a single slope (ramp) A/D converter.

At the beginning of a conversion cycle, the counter is reset and the ramp generator output is 0 V. At this point, the analog input is greater than the reference voltage and so the output of the comparator goes HIGH. This enables the clock to the counter and starts the ramp generator. Once the ramp voltage reaches the analog input, the ramp is reset and the binary count is stored in the latches by the control logic.

Single slope A/D converters are appropriate for very high accuracy of high-resolution (14-bit) measurements where the input signal bandwidth is relatively low. Besides the accuracy, these types of converters offer a low-cost alternative to others such as the successive-approximation approach. Typical applications of these converters are digital voltmeters, weighing scales, and process control. They are also found in battery-powered instrumentation due to their very low power consumption.

The accuracy of this converter depends on the precision of the ramp generator. In turn, the procision of the ramp generator depends on the *RC* time-constant of its integrator which varies with time and temperature. Any noise present at the comparator input while the ramp is near the threshold crossing can cause errors. So, this type of converter is more susceptible to noise.

Digital ramp A/D converter

As shown in Figure 2.18, this converter uses a binary counter as the register. A START pulse resets the counter and disables the AND gate. With all 0s at its input, the D/A converter's output is V_{ax} = 0 volts. Since $V_{ax} < V_{\text{in}}$, the opamp output is HIGH. When the START pulse returns LOW, the AND gate is enabled. As the counter advances, the D/A converter output, V_{ax}, increases one step at a time. This continues until V_{ax} reaches a step that just exceeds V_{in} by about V_T. The opamp output is then LOW disabling the AND gate. The A/D conversion is now complete and the contents of the counter are the digital representation of V_{in}. The digital data is lost at the next START pulse.

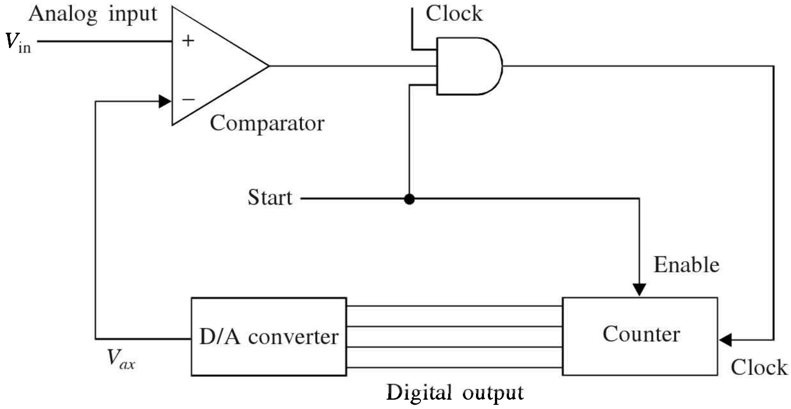


Figure 2.18 Block diagram of a digital ramp A/D converter.

Example 2.3

Given the following values for a digital ramp A/D converter

- Clock frequency = 100 kHz
- V_T = 100 μV
- The D/A converter's V_{ref} = 10.24 V and has a 10-bit input

determine (a) the digital equivalent representation for V_{in} = 3.728 V, (b) the resolution of the A/D converter, and (c) the conversion time required by this digital ramp A/D converter.

Solution

(a) The D/A converter has a 10-bit input and a 10.24 full-scale voltage output. So the possible steps are 2^{10} = 1024 and the step size is 10.24 V/1024 = 10 mV. Thus, V_{ax} increases in steps of 10 mV as the counter counts up from 0. Since V_{in} = 3.728 V and V_T = 0.1 mV, V_{ax} must reach 3.7281 V or more before the comparator switches LOW. This requires 3.7281V/10 mV which equals 372.81 or 373 steps. At the end of the conversion, then the counter holds the binary equivalent of 373, which is 0101110101. This is the desired digital equivalent of V_{in} = 3.728 V as produced by this A/D converter.

(b) The resolution of this A/D converter is equal to the step size of the D/A converter which is 10 mV. In per cent, it is (1/1024) × 100% ≅ 0.1%

(c) 373 steps are required to complete the conversion. Thus 373 clock pulses occurred at the rate of 1 per 10 μs giving a total conversion time of 3730 μs or 3.730 ms.

Dual slope A/D converter

The dual slope A/D converter operates on the principle of integrating the unknown input and then comparing the integration times with a reference cycle. The basic way is to use two (dual) slopes. Figure 2.19 shows the block diagram of this converter.

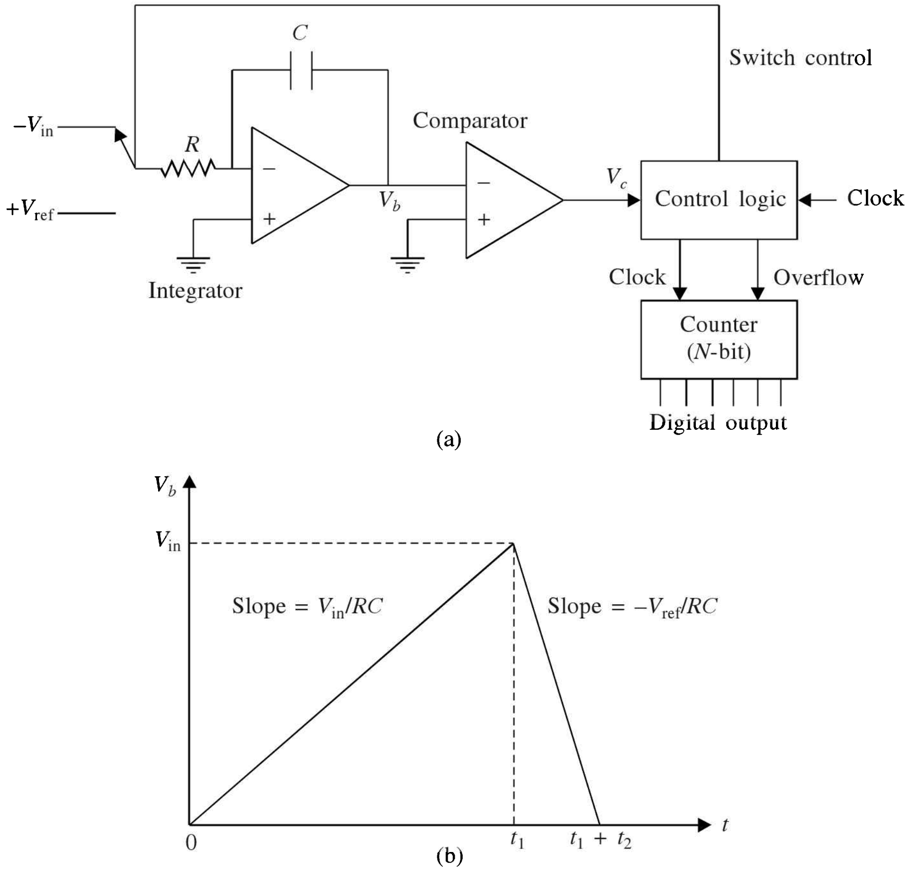


Figure 2.19 (a) Block diagram of dual slope A/D converter and (b) integrator output.

The counter is set to zero, and the analog input, $-V_{\text{in}}$, is connected to the integrator. The counter begins to increment and the output voltage from the integrator is given by

$$V_b = \int_0^{t_1} \frac{V_{\text{in}}}{RC}\,dt \qquad \text{(at a rate of } (V_{\text{in}}/RC) \text{ V/s)}$$

At t_1 the counter overflows, then the *Overflow* signal to the logic goes HIGH. At this point the integrator output is

$$V_b = \frac{V_{\text{in}} t_1}{RC} \qquad \text{(assuming } V_{\text{in}} = 0 \text{ at } t = 0\text{)}$$

The integrator is then connected to the reference voltage, V_{ref}, and the counter is reset. At t_1, the counter begins counting and the integrator output is given by

$$V_b = \int_{t_1}^{t_1 + t_2} -\frac{V_{\text{ref}}}{RC}\,dt \qquad \text{(at a rate of } (-V_{\text{ref}}/RC) \text{ V/s)}$$

At $t_1 + t_2$ the integrator output has changed by

$$V_b = \frac{V_{\text{ref}}\, t_2}{RC}$$

At this time, the integrator voltage is zero, the comparator output changes and the counter stops. Since the integrator output voltage is 0 V, we can equate the two values to give

$$V_b = \frac{V_{\text{ref}}\, t_2}{RC} = \frac{V_{\text{in}}\, t_1}{RC} \tag{2.1}$$

Since the counter overflows in the time t_1, we know that the counter counts to its maximum value 2^N and

$$t_1 = 2^N t_{\text{CLK}}$$

where t_{CLK} is the time period of the clock.

In t_2 the counter will count to a value n, then

$$t_2 = n t_{\text{CLK}}$$

Hence from Eq. (2.1)

$$\frac{V_{\text{ref}}\, n t_{\text{CLK}}}{RC} = \frac{V_{\text{in}}\, 2^N t_{\text{CLK}}}{RC}$$

and

$$n = \frac{V_{\text{in}} 2^N}{V_{\text{ref}}}$$

Thus n is proportional to the input voltage (rather the ratio of input and reference voltages) and the maximum count N but is independent of R and C.

Hence, the range of counter and reading the counter at $t_1 + t_2$ gives the desired digital output. The comparator determines when the integrator signal would reach zero. The control logic performs the necessary operations to control the switch depending upon the counter status and the comparator output. When the comparator switches, the counter is reset for the next integrate cycle and conversion.

Since this type of converter actually performs integration on the input signal, it nullifies the effects of noise (the integration process averages out the noise), provided the noise frequencies, f_{noise}, are such that

$$\frac{1}{f_{\text{noise}}} \geq t_1$$

where t_1 is the integration period. Thus the choice of t_1 determines what signals can be rejected, for example, the interference from the power supply mains.

Conversion time is determined by

$$t_{\text{conv}} = (2^N + n)t_{\text{CLK}} + t_{\text{D}}$$

where t_{D} is the time to zero the integrator.

The main benefits of this converter are the increased range, the increased accuracy and resolution, and the increased speed.

Successive approximation A/D converter

The conversion technique based on a Successive Approximation Register (SAR), also known as bit-weighting conversion, employs a comparator to weigh the applied input voltage against the output of an N-bit D/A converter (see Figure 2.20). Using the D/A converter output as a reference, this process approaches the final result as a sum of N weighting steps, in which each step is a single-bit conversion. Conversion time is directly proportional to conversion accuracy. This converter is faster than the dual slope A/D converter.

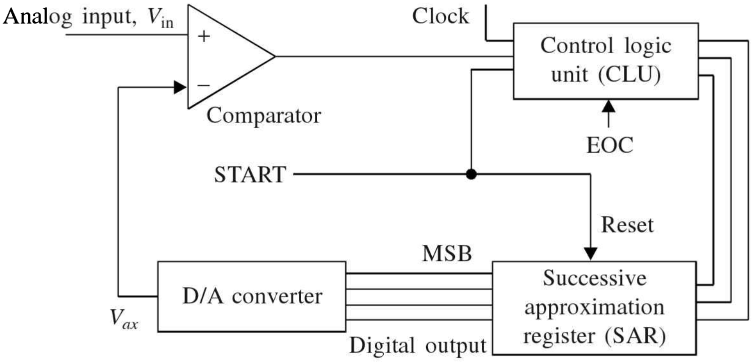


Figure 2.20 Successive approximation A/D converter.

The operation of the circuit is as follows: A START pulse clears all the bits, disables the Control Logic Unit (CLU) and sets End Of Conversion (EOC) output to 0. A/D conversion begins with the MSB = 1(1/2 of full-scale). The digital output of the register is converted to analog signal, V_{ax}, and compared with the analog input. If $V_{ax} > V_{\text{in}}$, then the last set bit is cleared back to 0 and the next lower significant bit to set to a 1. This process is repeated till all the bits have been checked. After all the bits have been checked, the EOC output is made a 1. The A/D conversion is now complete and the contents of the register are the digital representation of V_{in}. The digital data is lost at the next START pulse unless stored in some sort of memory device or location. The conversion is complete when all N bits have been compared. Thus, the conversion time is given by N/f_{CLK}, where f_{CLK} is the clock frequency.

Figure 2.21 shows the bit testing sequence for a 3-bit converter.

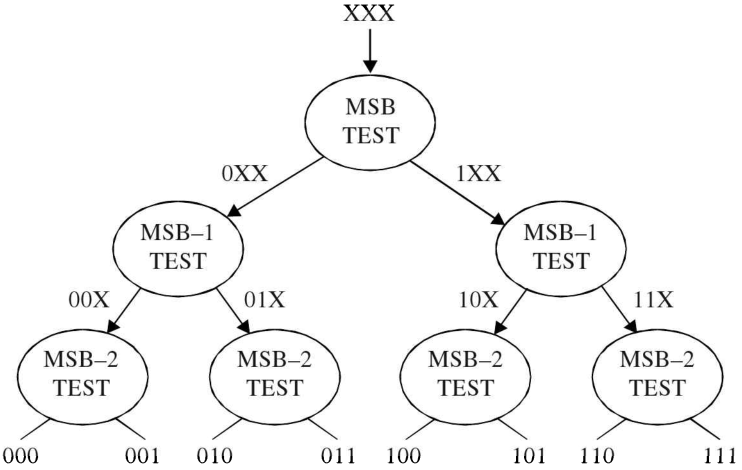


Figure 2.21 Bit testing sequence for a 3-bit converter.

In the above discussion we assumed that the analog input remains constant. If the input varies, during conversion time, the result will be wrong. Figure 2.22 shows the response of a 4-bit converter for a variable analog input. The final code depends upon the value of the input at the start of the conversion when the input is sampled. Therefore, we need a sample and hold circuit to meet this requirement.

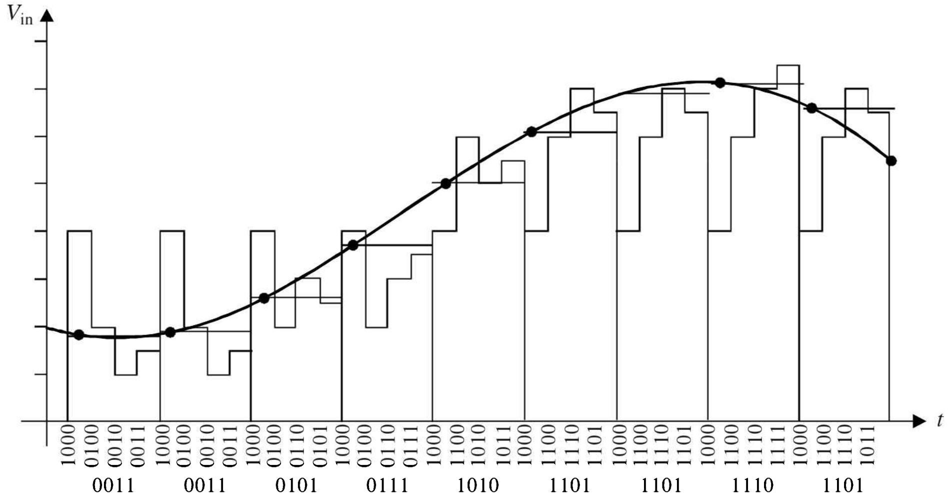


Figure 2.22 A 4-bit successive approximation A/D converter response for a variable analog input.

Successive approximation converters are more often used at lower speeds in higher-resolution applications. These converters are also well-suited for applications that have non-periodic inputs, since conversions can be started at will. This feature makes the architecture ideal for converting a series of time-independent signals.

Aliasing

The rate at which the samples of the input are taken is called *sampling rate* or *sampling frequency*. If the converter is being used to measure slow-changing signals, it could probably have a very slow sample frequency and still perform adequately. Conversely, if it is being used at high frequencies, the converter needs to be considerably faster.

Consider Figure 2.23 where analog input and digital output are plotted versus sampling time for a successive approximation A/D converter. Since the signal is now varying slowly the sample rate is more than adequate to capture the general trend of the signal. Now consider the signal shown in Figure 2.24, where the input changes fast, but is being sampled at the same time intervals as in Figure 2.23.

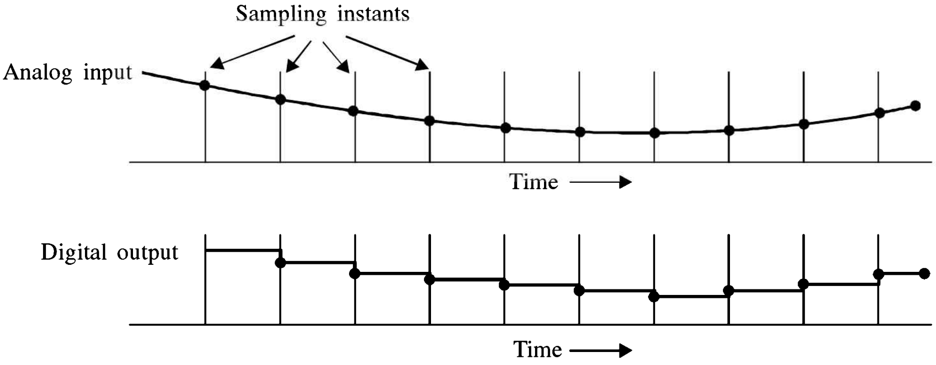


Figure 2.23 Analog input and digital output versus sampling time.

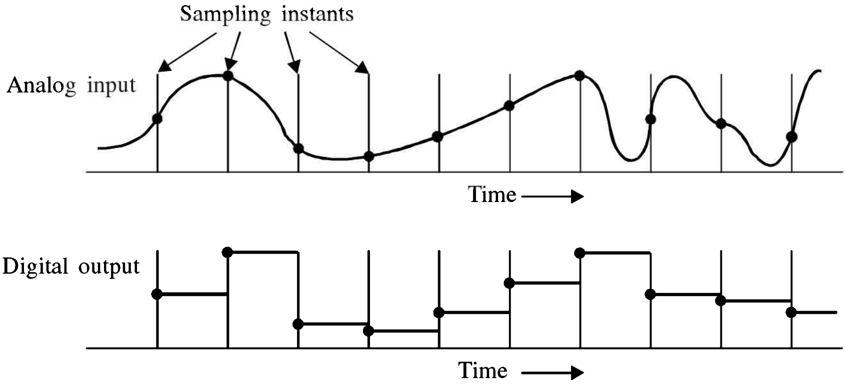


Figure 2.24 Sampling of fast varying signal.

Note also how in the latter portions of the analog signal, the digital output fails to reproduce the true shape. Even in the first section of the analog waveform, the digital reproduction deviates substantially from the true shape of the wave. This is because when the sample period is too long, substantial details of the analog signal are missed out. Therefore, the sample interval should be small enough to capture essential changes in the analog waveform. The sampling frequency is related to *Nyquist frequency* according to which it must be equal to twice the maximum frequency of the input signal. For example, if an A/D converter circuit has a sample frequency of 5 kHz, the highest-frequency waveform it can successfully resolve is the Nyquist frequency of 2.5 kHz.

If an A/D converter is subjected to an analog input signal whose frequency exceeds the Nyquist frequency for that A/D converter, the converter will output a digitized signal of falsely low frequency. This phenomenon is known as aliasing (see Figure 2.25).

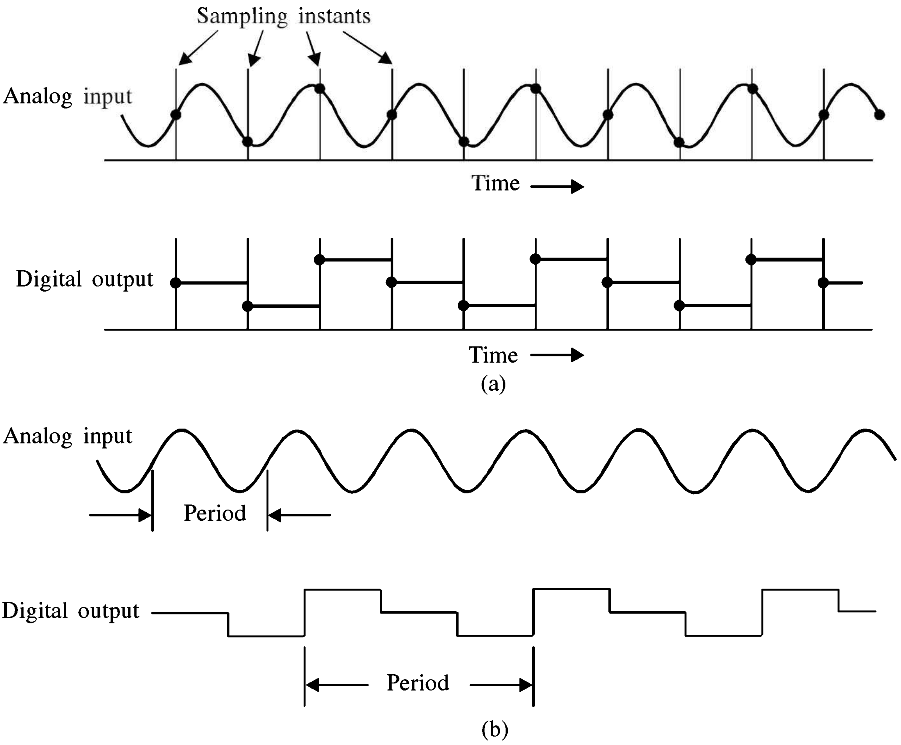


Figure 2.25 Aliasing and its effect on the output wave shape.

We notice from Figure 2.25(b) that the period of the output waveform is much longer (slower) than that of the input wave form, and how the two wave shapes are not at all similar. It should be understood that the Nyquist frequency is an absolute maximum frequency limit for an A/D converter, and does not represent the highest practical frequency measurable. To be safe, we should not expect an A/D converter to successfully resolve any frequency greater than one-fifth to one-tenth of its sample frequency.

A practical means of preventing aliasing is to place a low-pass filter (anti-alias filter) before the input of the A/D converter, to block any signal frequencies higher than the practical limit. This way, the A/D converter circuitry is prevented from getting any higher frequencies and thus will not try to digitize them. It is generally considered better that such frequencies go unconverted than to have them be *aliased* and appear in the output as false signals.

The commonly used filters are Butterworth filter, Chebyshev filter, Inverse Chebyshev filter, Cauer filter and Bessel-Thomson filter. Each filter has its own advantages and disadvantages. We will not discuss the details of these filters here.

Undersampling

For high frequency input, we need a high sampling frequency. But many A/D converters may not be able to operate at this high frequency. Undersampling is a powerful tool that can be used effectively in selected applications. It allows an A/D converter to behave like a mixer in that it can take a high-frequency signal and create an image that is lower in frequency. This method effectively uses the A/D converter as a downconverter by shifting higher bandwidth signals into the A/D converter's desired band of interest. Another key advantage is that it allows us to use an A/D converter with a sample rate that is lower than Nyquist causing aliasing, which usually has significant cost advantages. With proper filtering of the input signal and proper frequency selection, the aliased components at lower frequency are selected and converted.

One drawback of undersampling is that unwanted signals can appear in the desired band of interest and we cannot differentiate them from the desired signal. Also, when undersampling, the frequency range at the A/D converter input is often very wide.

Sample and hold circuits

In most A/D converters, it is difficult to ascertain when the digital output matched the analog input signal, unless it was static. If the conversion is required at a specific instance in time, then a sample and hold circuit is used to capture the signal value and hold it whilst the conversion takes place. If the analog input signal changes, then it will not affect the result. The basic sample and hold circuit comprises a switch, S, and a capacitor, *C*, as shown in Figure 2.26.

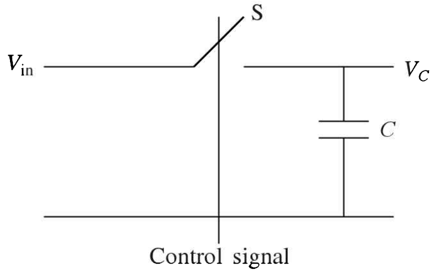


Figure 2.26 Basic structure of a sample and hold circuit.

When the switch is closed, the voltage across the capacitor, V_C, follows or tracks the applied voltage, V_{in}. When the switch is opened, the capacitor retains the analog voltage applied at that instant of time.

Practically, the signal is buffered, as illustrated in Figure 2.27, to minimize the effects of loading. Also, a field effect transistor (FET) is used as the switch since it operates like an ideal switch, with an 'on' resistance <100 Ω and an 'off' resistance > 10^8 Ω with zero offset voltage.

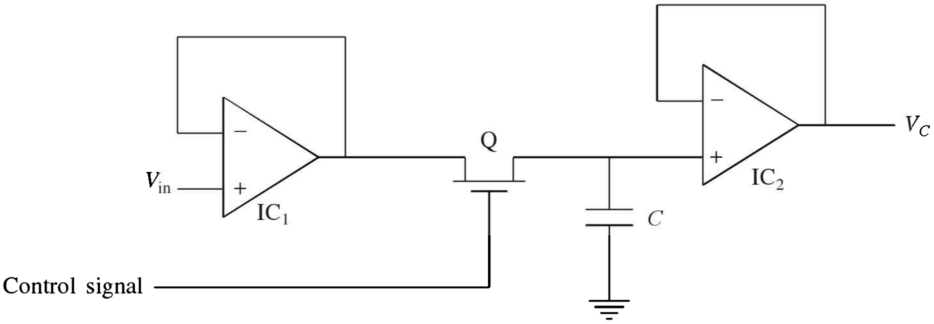


Figure 2.27 A practical sample and hold circuit.

The 'hold' time depends upon the characteristics of the capacitor, since its charge leaks over time. As a result the voltage across the capacitor, V_C, decays or droops. Now, we move on to the response parameters of sample and hold circuits (see Figure 2.28).

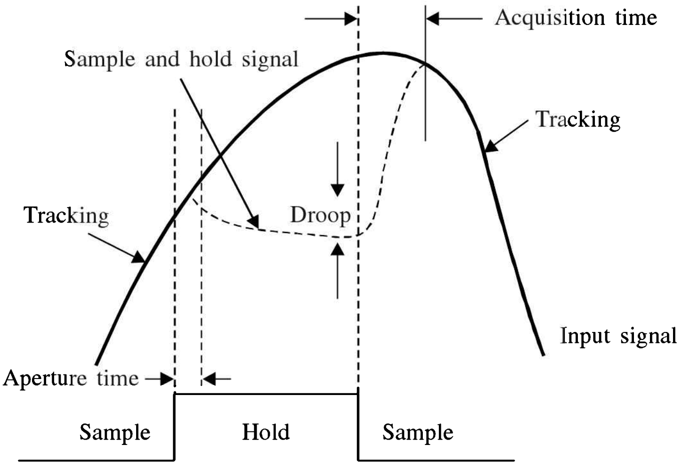


Figure 2.28 Response parameters of sample and hold circuits.

Response parameters

Aperture time. Time required for the switch to open. This is also called aperture uncertainity.

Acquisition time. Time it takes the output of the sample and hold circuit to 'track' the input signal within a specified accuracy. It depends upon the maximum charging current that the input amplifier can supply to the capacitor plus the switch operation time.

Droop rate. The rate of change of the output voltage when the circuit is in the 'hold' mode, and is a function of the leakage or droop current, i_d, and the capacitance, C.

$$\frac{dV_C}{dt} = \frac{i_d}{C}$$

Example 2.4

The maximum charging current of the input opamp in a sample and hold circuit is 50 mA and the droop current is 100 pA. What is the expected droop rate if the acquisition time for a 3 V step is 6 μs?

Solution

The value of C is given by

$$C = \frac{i_C t}{V_C} = \frac{50\,(\text{mA}) \times 6\,(\mu\text{s})}{3\,(\text{V})} = 0.1\ \mu\text{F}$$

The droop rate is

$$\frac{dV_C}{dt} = \frac{i_d}{C} = \frac{100\ (\text{pA})}{0.1\,(\mu\text{F})} = 1\ \text{mV/s}$$

Considerations for choosing C

- C should be large enough to minimize the 'droop' caused by leakage currents in Q and IC2
- C should be small enough to track fast signals since it forms a low-pass filter with the ON resistance of Q.

Parallel or flash A/D converter

In the flash A/D converter the analog input is compared with many reference voltages simultaneously. For an n-bit converter, $2^n - 1$ analog comparators are used. Various reference voltages are generated by a resistive divider of 2^n resistors connected between two voltage sources (or a single voltage source) representing the lower and upper limits of the desired conversion range. The basic structure for a 3-bit converter is shown in Figure 2.29.

The comparators generate $2^n - 1$ bits of information in which all of the comparators below the input voltage produce 1s, and all of the comparators above the input voltage produce 0s. A priority encoder is used to convert the comparators output to the proper binary output. Table 2.2 gives the truth table of the priority encoder.

The biggest advantage of this converter is its high speed and it usually operates in the 100-MHz to several-GHz range, providing between 6 and 10 bits of resolution. The major disadvantage is the circuit complexity. Because of the complexity and the heavy reliance on perfectly matched components, these are available mostly in the single chip form. For each additional bit of resolution, the number of comparators double, which drives up both chip area and power consumption.

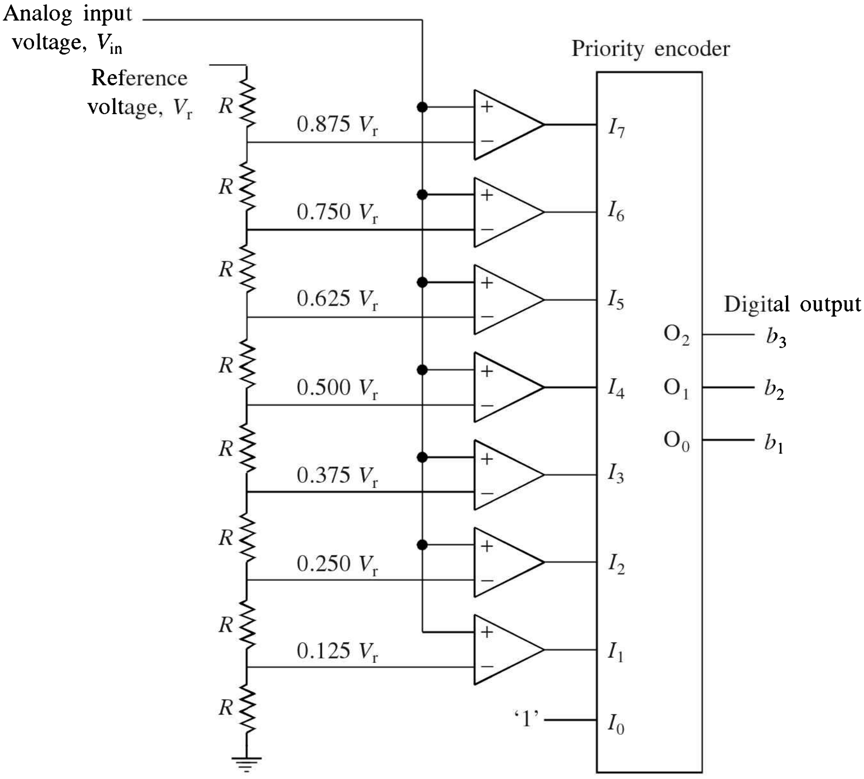


Figure 2.29 A 3-bit flash A/D converter.

Table 2.2 Truth table of priority encoder

Input								*Output*		
I_0	I_1	I_2	I_3	I_4	I_5	I_6	I_7	O_2	O_1	O_0
1	0	0	0	0	0	0	0	0	0	0
×	1	0	0	0	0	0	0	0	0	1
×	×	1	0	0	0	0	0	0	1	0
×	×	×	1	0	0	0	0	0	1	1
×	×	×	×	1	0	0	0	1	0	0
×	×	×	×	×	1	0	0	1	0	1
×	×	×	×	×	×	1	0	1	1	0
×	×	×	×	×	×	×	1	1	1	1

Sigma delta A/D converter

Conventional converters, as illustrated in Figure 2.30, are often difficult to implement in fine-line VLSI technology with reasonably low power consumption. These difficulties arise because conventional methods need analog components that are precise and highly immune to

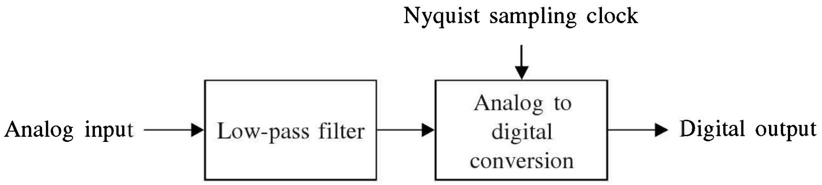


Figure 2.30 Conventional Nyquist A/D converter.

noise and interference. However, the conversion rate of conventional converters is Nyquist rate, i.e. sampling frequency is twice the signal bandwidth. That is why conventional converters are usually referred to as Nyquist converters.

The low-pass filter at the input attenuates high frequency noise and out-of-band signals to prevent them from aliasing into the desired signal when sampled. One problem with low-pass filters is an anomaly called *ripple*. In order to allow frequencies at Nyquist to pass, the filter must open and close very fast. This filter is not fast enough which allows some frequencies above Nyquist to pass through, which is ripple. Symptoms of ripple include ringing and phase distortion.

The analog-to-digital circuit can take a number of different forms such as flash converters for fast operation, successive-approximation converters for moderate rates, and ramp converters for high resolution. All of them are called *multibit converters.*

However, there are techniques that can be used to achieve higher resolution measurements and signal-to-noise ratio (SNR). These are called low-bit converters. Low-bit conversion divides the signal in terms of time, keeping the amplitude changes constant. Figure 2.31 shows the concept behind low-bit conversion and its relation to multibit conversion. When the amplitude representations in Figure 2.31(a) are placed on their side as shown in Figure 2.31(b), the signal is represented in terms of time with 1 bit. When these signals are low pass filtered, the result is an analog signal just as is obtained in the multibit method. However, since the weighted bits of the binary sample are transformed into a low-bit data stream using digital methods, the major sources of nonlinearity are avoided.

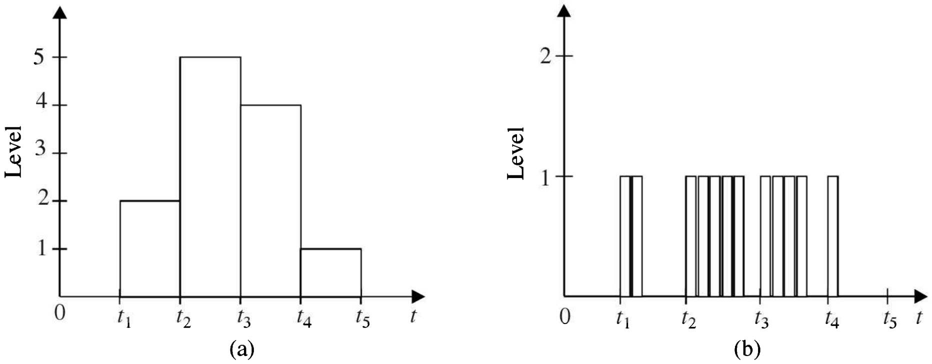


Figure 2.31 (a) Multibit representation and (b) low bit representation.

Low-bit conversion techniques use oversampling and digital signal processing algorithms like noise shaping and decimation. These low-bit converters normally operate with one bit and, therefore, are also called 1-bit converters.

Oversampling

Oversampling moves the filtering requirements from the analog domain to the digital domain and provides a less expensive yet better way to get good quality encoding without the need for expensive analog low-pass filters.

The required sampling frequency in accordance with the Nyquist theorem is the Nyquist frequency, $f_n = 2f_B$, where f_B is the highest frequency component of interest in the input signal. Sampling frequencies f_s above f_n is *oversampling*, and will increase the resolution of a measurement and signal-to-noise ratio.

When a quantized signal is sampled at frequency f_s, all of its power folds into the frequency band $0 < f < f_s$. Then, if the quantization noise is white, the spectral density of the sampled noise is given by

$$E(f) = \sqrt{\frac{e^2_{\text{rms}}}{f_s/2}} = e_{\text{rms}}\sqrt{\frac{2}{f_s}}$$

We have seen that, assuming equal steps, the total error (mean square quantization noise) is

$$e^2_{\text{rms}} = \frac{q^2}{12}$$

where q is the value of one step.

We can use this result to analyze oversampling modulators. Consider a signal lying in the frequency band $0 \le f \le f_B$. The oversampling ratio (OSR), defined as the ratio of the sampling frequency f_s to the Nyquist frequency $2f_B$, is given by the integer

$$\text{OSR} = \frac{f_s}{2f_B}$$

The in-band noise power is given by the equation

$$n_0^2 = \int_0^{f_B} E^2(f)\, df = e^2_{\text{rms}}\left(\frac{2f_B}{f_s}\right) = \frac{q^2}{12.\text{OSR}} = \frac{1}{12.\text{OSR}}\left(\frac{V_{\text{ref}}}{2^n}\right)^2 = \frac{(\text{Resolution})^2}{12.\text{OSR}}$$

where n_0 is the noise power output, n is the number of A/D converter bits and V_{ref} is the reference voltage.

$$\text{Noise level as a function of frequency} = n_0 = \frac{q}{\sqrt{12.\text{OSR}}}$$

The signal-to-noise ratio, SNR, is given by

$$\text{SNR (dB)} = 6.02n \text{ dB} + 1.76 \text{ dB} + 10 \log (\text{OSR})$$

Thus, we see that oversampling reduces the in-band quantization noise from ordinary quantization by the square root of the oversampling ratio. Therefore, each doubling of the sampling frequency decreases the in-band noise by 3 dB and thus increases the resolution by half a bit.

For each additional bit of resolution or 6 dB of noise reduction, we oversample by a factor of four. In general,

$$f_{s2} = 4^{w} f_{s1}$$

where w is the number of additional bits of resolution desired, f_{s1} is the original sampling frequency requirement, and f_{s2} is the oversampling frequency.

If we are using the 12-bit on-chip ADC and wish to have the accuracy of a 16-bit ADC, we need an additional 4 bits of resolution. 4^4 is 256. Thus, we need to oversample by a factor of 256 times the Nyquist rate. If the desired signal is band-limited to 50 Hz (f_B = 50 Hz), then we must oversample at 100 Hz $*$ 256 = 25.6 kHz. We improve the effective resolution by improving the SNR in our frequency band of interest.

Figure 2.32 shows the power spectral density, $E(f)$, of the quantization noise for Nyquist rate sampling with rate f_{s1} and oversampling rate f_{s2}. For Nyquist rate sampling where the signal band, $f_B = f_{s1}/2$, all the quantization noise power, represented by the area of the tall shaded rectangle, occurs across the signal bandwidth. In the oversampled case, the same noise power, represented by the area of the unshaded rectangle has been spread over a bandwidth equal to the sampling frequency, f_{s2}, which is much greater than the signal bandwidth, f_B. Only a relatively small fraction of the total noise power falls in the band $[-f_B, f_B]$, and the noise power outside the signal band can be greatly attenuated with a digital low-pass filter following the A/D converter.

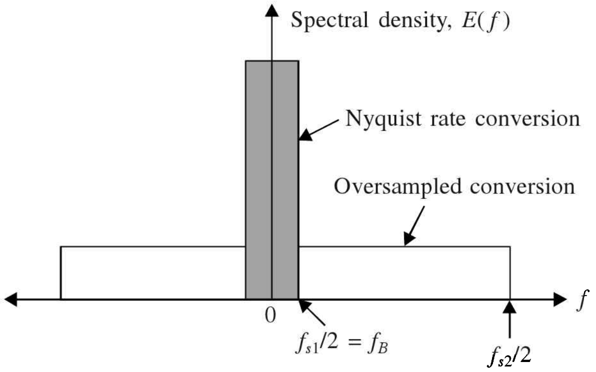


Figure 2.32 Quantization noise for Nyquist rate and oversampled conversion.

Sigma delta modulator

The sigma delta (or delta sigma) modulator quantizes the difference (delta) between the current signal and the sum (sigma) of previous differences. A first-order (single integration) sigma delta modulation encoder is shown in Figure 2.33. The input to the quantizer is the

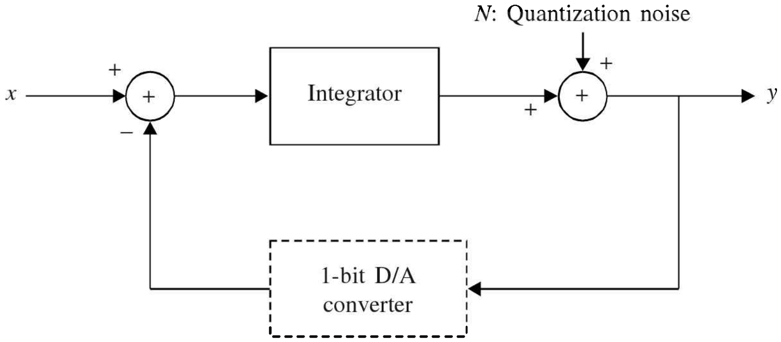


Figure 2.33 Sigma delta modulation encoder.

integral of the difference between the input and the quantized output. The quantized output is normally a single bit output, which is generated by using a comparator. This single bit output is again converted to analog equivalent by using a 1 bit D/A converter, which may simply be a capacitor that charges and discharges, depending upon the value of the single bit.

In the frequency (s) domain, integrator (sigma function) is represented as $1/s$. The signal transfer function when noise, N, is zero is given by the equation

$$y = (x - y)\frac{1}{s}$$

or

$$H_x = \frac{y}{x} = \frac{1/s}{1 + (1/s)} = \frac{1}{s+1}\text{: Low-pass filter (see the response in Figure 2.34(a)).}$$

The noise transfer function when signal, x, is zero, is given by the equation

$$y = -y\frac{1}{s} + N$$

or

$$H_n = \frac{y}{N} = \frac{1}{1 + (1/s)} = \frac{s}{s+1}\text{: High-pass filter (see the response in Figure 2.34(b)).}$$

It can be seen from Figure 2.34 as the difference between the input signal and the fed back signal loops, it low pass filters the signal and high pass filters the noise. If the system is

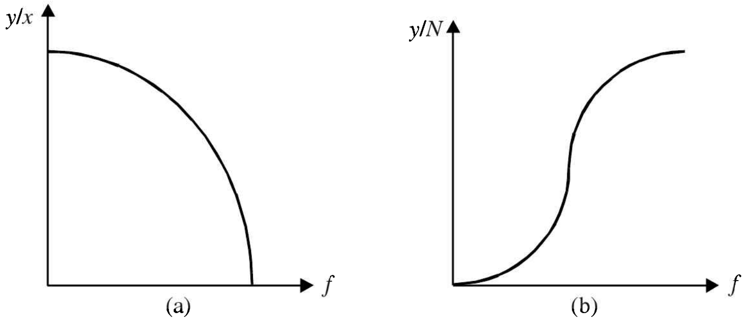


Figure 2.34 (a) Response of the signal transfer function and (b) response of the noise transfer function.

designed so that the signal's frequency content is less than the filter's cut-off frequency, the signal will not be affected. However, oversampling spreads the noise over a wider bandwidth; hence the noise is reduced in the desired frequency band. This is also called *noise shaping*. Figure 2.35 shows how the noise varies with the order of the noise shaping circuit. The higher the order (multiple integration) of noise shaping, the lower the noise in the desired band of frequency. But beyond this frequency, the power of the noise increases. Higher-order circuits need special design techniques to ensure stability.

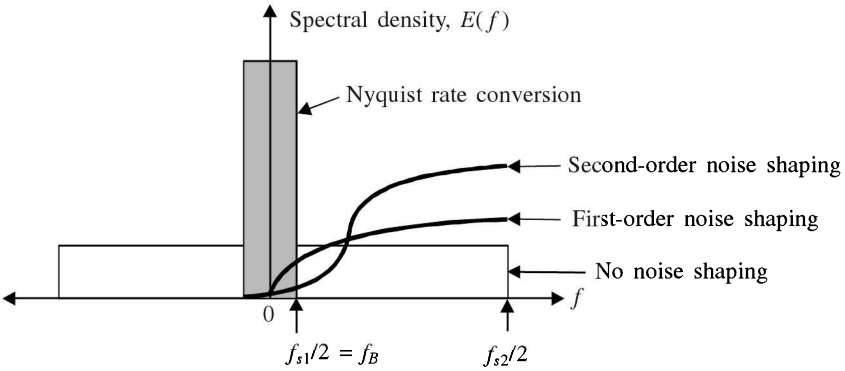


Figure 2.35 Quantization noise for Nyquist rate and oversampled conversion with different orders of noise shaping (negative frequency not shown).

For first-order noise shaping, the in-band noise power is given by the equation

$$n_0^2 = \int_0^{f_B} E^2(f) H_n^2 \, df$$

$$\propto \left(\frac{1}{\text{OSR}}\right)^3$$

The maximum signal-to-noise ratio, SNR, is given by

$$\begin{aligned}\text{SNR (dB)} &= 6.02n \text{ dB} + 1.76 \text{ dB} - 5.17 \text{ dB} + 30 \log (\text{OSR}) \\ &= 6.02n \text{ dB} - 3.41 \text{ dB} + 30 \log (\text{OSR})\end{aligned}$$

Therefore, each doubling of the sampling frequency decreases the in-band noise by 9 dB and thus increases the resolution by 1.5 bits.

For second-order noise shaping, the maximum signal to noise ratio, SNR, is given by

$$\begin{aligned}\text{SNR (dB)} &= 6.02n \text{ dB} + 1.76 \text{ dB} - 12.9 \text{ dB} + 50 \log (\text{OSR}) \\ &= 6.02n \text{ dB} - 11.14 \text{ dB} + 50 \log (\text{OSR})\end{aligned}$$

Each doubling of the sampling frequency decreases the in-band noise by 15 dB and thus increases the resolution by 2.5 bits.

A general formula of the maximum signal to noise ratio with *k*th-order noise shaping is

$$\text{SNR} = 6.02n \text{ dB} + 1.76 \text{ dB} - 10 \log \frac{2k+1}{\pi^4} + (2k + 1)\ 10 \log (\text{OSR}\downarrow)$$

Decimation

The sigma delta modulator output is applied to the digital decimator. The basic tasks of the digital decimation process are:

(a) To remove the shaped out-of-band quantization noise. The low-pass filtering in the frequency domain constitutes the averaging process.

(b) To reduce the sampling rate (undersampling) to the Nyquist rate which minimizes the amount of information for subsequent transmission, storage, or digital signal processing.

(c) Anti-aliasing (because the signals are seldom completely band-limited in practice)

Figure 2.36 shows the block diagram of a decimation filter.

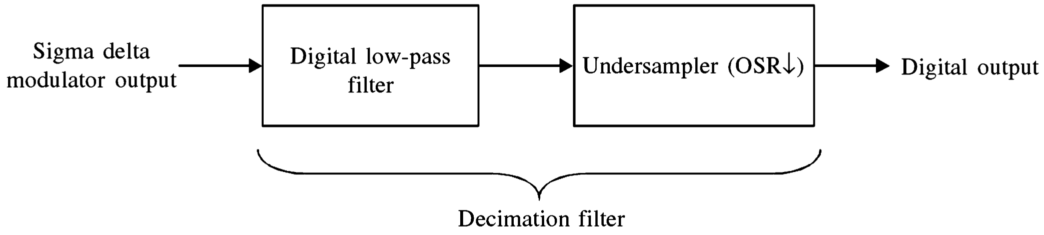


Figure 2.36 Block diagram of a decimation filter.

The undersampler (OSR) samples the filter output at the original (without oversampling) signal Nyquist rate.

Decimation can be described through a simple example. Sixteen 1 bit values can be reduced through a 16:1 decimation to a single low bit value; for example, values 1, 0, 1, 0, 0, 1, 0, 0, 1, 0, 1, 1, 1, 1, 0, 0 would be decimated to 8/16, or 0.5. Since there is only one (low bit) output value for every sixteen input values, the decimator has decreased the sample rate by 16:1. It means decimation replaces 1-bit coding with 16-bit coding while reducing the sampling rate. For example, it provides a lower sampling rate and generates a pulse code modulated (PCM) output.

Combining all the submodules, Figure 2.37 shows the block diagram of a sigma delta A/D converter.

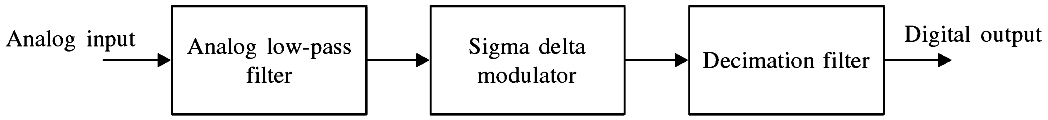


Figure 2.37 Block diagram of a sigma delta A/D converter.

Characteristics of a sigma delta A/D converter

The general characteristics of the sigma delta converter are:

- High resolution
- Low in-band noise
- Low bit internal quantization (1 or 2 bit)
- Inherent linearity and high tolerance to circuit imperfection
- Simple circuitry and ease of VLSI implementation (i.e. low implementation costs)
- Baseband or band-pass conversion
- Easy realization of anti-aliasing filters
- Increased programmability in digital domain

Performance characteristics of A/D converters

Code width. The width of a given output code is the range of analog input voltages for which that code is produced. The code widths are referenced to the weight of 1 LSB.

$$1\ \text{LSB} = \frac{V_r}{2^N}$$

where V_r is the reference (full scale voltage) and N is the number of bits in the output. All codes will have a 1 LSB for an ideal A/D converter.

Resolution. The resolution is specified in bits and determines how many distinct output codes (2^N) the converter is capable of producing. For example, a 10-bit A/D converter produces $2^{10} = 1024$ output codes.

Dynamic range. The digitization of an analog voltage signal is specified over a particular range. That is to say, a maximum and minimum input voltage is defined over which the quantization should occur. The dynamic range in dB for a digital value can be expressed as

$$\text{Dynamic range in dB} = 20 \log 2^N = 6 \times (\text{number of bits})$$

Accuracy. Accuracy tells us how close is the digital output to the theoretically expected digital output. It is a function of the internal circuitry and noise from the external sources connected to the input. Accuracy is generally specified as

$$\underbrace{\pm\ 0.01\%\ \text{FSR}}_{\text{Analog inaccuracies}} \qquad \underbrace{\pm\ 1/2\ \text{LSB}}_{\text{Quantization error}}$$

where FSR = full scale reading and LSB = least significant bit. The first term includes all the analog sources of errors like linearity, drift and component tolerances. It also includes the *offset* and *gain* errors.

Offset error. Offset error is defined as a deviation of the code transition points (analog input voltages at which the output code switches from one code to the other) that is present across all output codes. It is measured by finding the difference between the actual location of the first code transition and the desired location of the first transition. Offset error can be positive or negative. This is expressed in % FSR (see Figure 2.38).

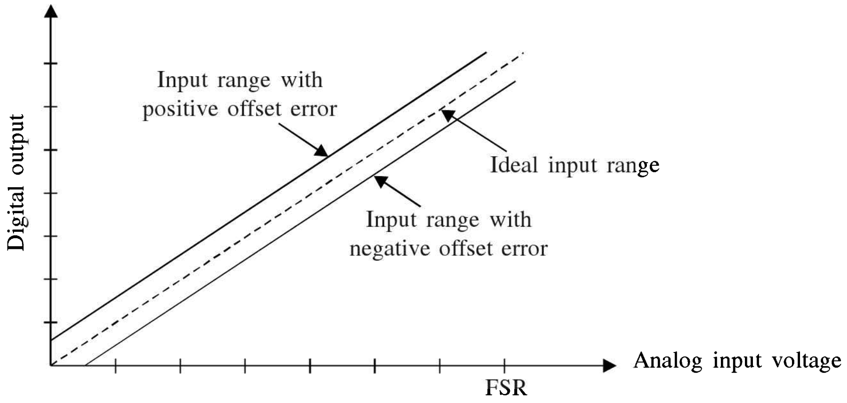

Digital output
Input range with positive offset error
Ideal input range
Input range with negative offset error
Analog input voltage
FSR

Figure 2.38 Offset error and its effect on the unipolar input range.

Related to the offset error is the offset error temperature coefficient, which is the change in offset over temperature. This is usually specified in ppm/°C.

Gain error. The gain error determines the amount of deviation from the ideal slope of transfer characteristics of the A/D converter. It can be positive or negative. It is expressed in % FSR (see Figure 2.39).

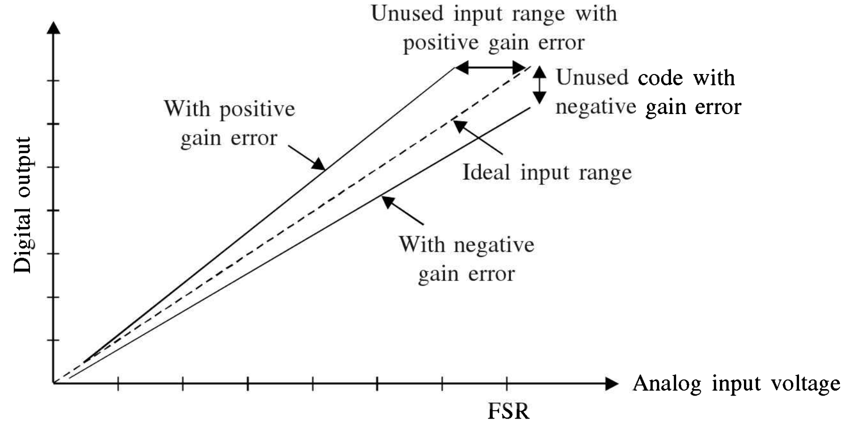
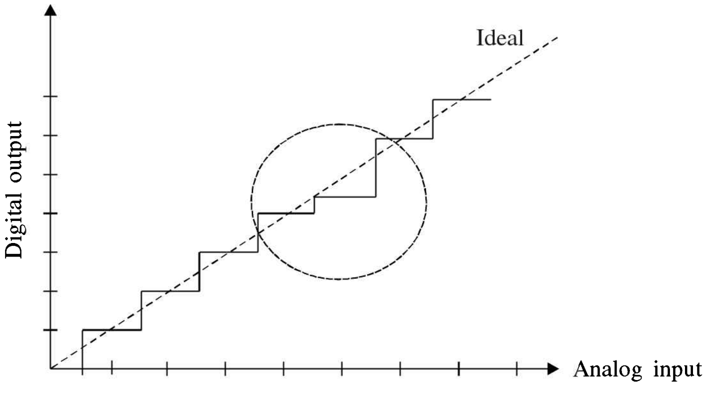


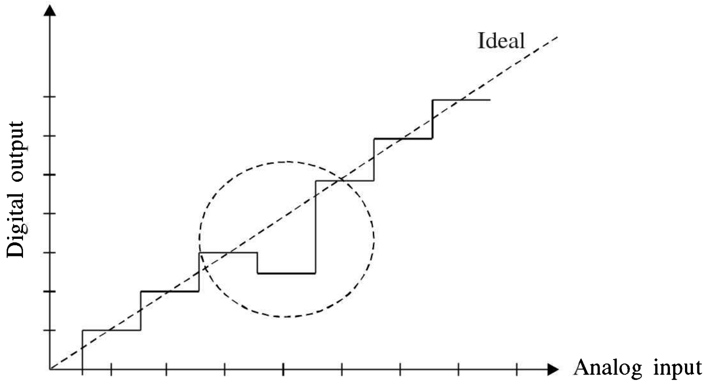
Figure 2.39 Gain error and its effects on the unipolar input range and the code.

Related to the gain error is the gain error temperature coefficient, which is the change in gain over temperature. This is usually specified in ppm/°C.

The offset and gain errors still keep the transfer characteristics linear. The characteristics, however, may not be linear and may not always increase (monotonic) with the increase in the input (see Figure 2.40).



(a)



(b)

Figure 2.40 (a) Nonlinearity and (b) non-monotonocity.

Following types of error values are also given by the manufacturers:

Differential nonlinearity (DNL) error

The differential nonlinearity error shown in Figure 2.41 (sometimes seen as simply differential linearity) is the difference between an actual step width and the ideal value of 1 LSB. Therefore if the step width is exactly 1 LSB, then the differential nonlinearity error is zero. If the DNL exceeds 1 LSB, there is a possibility that the converter can become non-monotonic. This means that the magnitude of the output gets smaller for an increase in the magnitude of the input. There is also a possibility that there can be missing codes, i.e. one or more of the

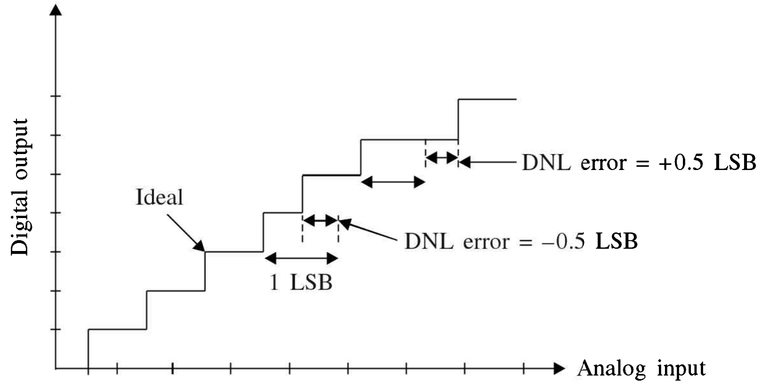


Figure 2.41 Differential nonlinearity error.

possible 2^n binary codes are never output. As an example, when using an A/D converter, as the analog input voltage increases, if the next code transition occurs at a voltage 1.5 LSB away from the previous transition, the DNL error is 0.5 LSB; if it occurs at 0.5 LSB away, the DNL error is −0.5 LSB.

We can define DNL error as follows:

$$\% \text{ DNL error} = \frac{\pm(V_{\text{in}} - V_{\text{LSB}})}{V_{\text{LSB}}} \times 100\%$$

Sometimes the DNL error is expressed as % of FSR.

Integral nonlinearity (INL) error

The integral nonlinearity error shown in Figure 2.42 (sometimes seen as simply linearity error) is the deviation of the values on the actual transfer function from a straight line. The deviations are measured at the transitions of each step. This straight line can be either a best straight line which is drawn so as to minimize these deviations or it can be a line drawn

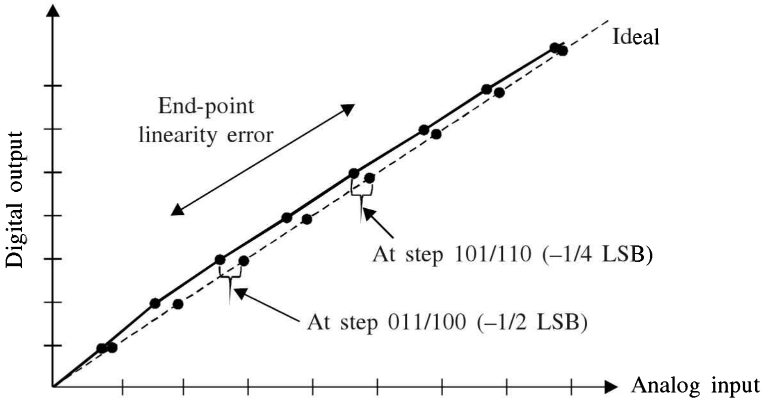


Figure 2.42 Integral nonlinearity error.

between the end-points of the transfer function once the gain and offset errors have been nullified. The second method is called the end-point linearity and is the usual definition adopted since it can be verified more directly.

The name integral nonlinearity derives from the fact that the summation of the differential nonlinearities from the bottom up to a particular step, determines the value of the integral nonlinearity at that step.

$$\text{INL error} = \Sigma\, [V_{\text{in}}(i) - V_{\text{LSB}}] \text{ for } i = 0 \text{ up to the given steps}$$

Sometimes it is expressed as % of FSR.

Both the INL and DNL errors affect ac applications as distortion and spectral harmonics (spurs). In dc applications, they result in an error in the dc output voltage.

Absolute error

Absolute error is the sum of the offset, gain and integral nonlinearity errors. In other words, this is the amount of deviation from the ideal A/D transfer characteristics without compensating for gain or offset errors. The absolute error is also called *total unadjusted error*.

Settling time

This is the time taken between the application of a full scale input step and the stabilization of the signal within an error band.

Conversion time

This is the time required to obtain the digital output after the analog input is disconnected from the holding capacitor of sample and hold circuit. It is generally specified in A/D clock cycles. The minimum period of the clock is specified to obtain specific accuracy for the A/D converter.

Power supply rejection ratio

The power supply rejection ratio is sometimes called *power supply sensitivity*. It is the ability of the converter to reject ripple and noise on its power inputs. DC applications may not be adversely affected. Poor power supply rejection can cause spurs and harmonic distortion in ac applications, as external frequency components leak into the output and modulate with it.

It is defined as a per cent of the full scale reading change in the output to per cent of change in either positive, negative or digital supply voltage about the nominal power supply voltages.

Drift

Many A/D converter parameters like gain and various types of errors vary with temperature. This is also called *drift* and is specified as ppm of FSR/°C with reference to 25°C temperature.

2.4 DIGITAL METER DISPLAYS

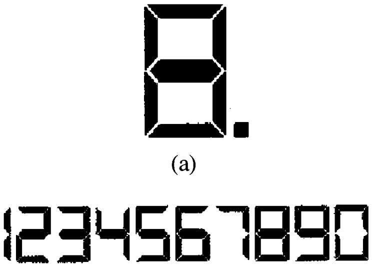
The DMM typically offers more functions, better accuracy, greater ease of reading, and greater reliability than do many analog meters. Analog meters have at least one important advantage over the digital meter. They can respond to transient (fast and brief) changes and trends in a measured quantity better than most digital meters because DMMs are too slow to react to these changes.

There are two basic types of displays used in DMMs—LCD (liquid-crystal display) and LED (light-emitting diode) readouts. The amount of reflected light from the LCD depends on the current flowing through it.

In an LED readout, when biased in the forward direction, a p–n junction diode emits light as a result of recombination of excess electron-hole pairs. The average forward drop across an LED is about 1.2 V. The typical ohmmeter will not check most LEDs because it requires about 1.2 V to forward bias an LED and most ohmmeters do not have that much terminal voltage. As a result most ohmmeters show an LED open in both directions.

The LCD display is most popular for battery-powered meters because they require much less operating current than the LED types. Most DMMs operate on 9 V battery with life from a few hundred hours to 2000 hours. They are very difficult to see when lighting is poor or absent entirely. The LED display, however, can be seen in the dark and is faster to react to changes in electrical values than the LCD displays.

Both the LED and LCD displays are in a seven-segment format (Figure 2.43). In other words, each digit in the display has seven separate segments or sections. Each segment is independently controlled. Numbers are formed by lighting the appropriate segments that will outline the number. In addition to the seven segments, there is also a decimal point. Digital decoders and drivers are used to interface digital output from A/D converters to the display.



(a)

(b)

Figure 2.43 Seven-segment display: (a) basic display and (b) numbers display.

2.5 SPECIFICATIONS OF DIGITAL METERS

2.5.1 Display Digits and Counts

Traditionally, DMM displays have been specified as “3-1/2-digit”, for example. This means that there are three complete digits, each capable of displaying the numbers zero to nine, and one additional preceding digit which may display only a zero (may be blanked) or a one. Therefore it will display full-scale reading of 1999.

The latest DMMs show the full-scale range, say, from 3999 to 39999 or more. These are dubbed as 3-3/4- and 4-3/4-digits respectively. This description is even less intuitive than it was for 3-1/2-digits. Now the fractional digits are specified by the number of "counts" that may be displayed. For example, the 3-1/2-digit display is described as 2000 count (1999 plus the reading of 0). From that description, it becomes readily obvious what the display is capable of showing. 3-3/4-digits becomes 4000 counts and 4-3/4-digits becomes 40000 counts. Table 2.3 shows the relationship between digits and counts for the more common DMM displays.

Table 2.3 Digits vs. counts

Digits	*Counts*
3-1/2	2000
3-3/4	4000
4-1/2	20000
4-3/4	40000
4-4/5	50000

The number of counts usually applies to the dc volts function. Fewer counts may be displayed on the same instrument for certain other functions. For instance, a 40000 count DMM may be limited to 4000 counts when measuring capacitance.

2.5.2 Resolution

Resolution is a measure of the smallest increment that may be discerned. By knowing the resolution of a meter, we can determine if it is possible to see a small change in the measured signal. For example, if the DMM has a resolution of 1 mV on the 4 V range, it is possible to see a change of 1 mV while reading 1 V.

For certain measurements, low count meters may be more useful than high count meters. For example, 4000 count meters offer better resolution than a 2000 count meter which would not be able to measure down to a tenth of a volt if we were to measure 200 volts or more. However, a 4000 count meter will display a tenth of a volt up to 400 volts. This is the same resolution as the more expensive 20000 count meter will give until we exceed 400 volts.

Further, A/D converter resolution may also limit the resolution. For example, at first glance, it seems that 10.000 volts measured with a 40000 count DMM would read to a resolution of 0.001 volt. This is usually the case where the DMM's A/D converter resolution exceeds that of the display but some have less resolution than the display. In this case, the last digit could read 0, 1, 2, 3, 5, 6, 7, 9, etc. with a linearly increasing voltage. Note that only eight out of the ten possible values were displayed. This is an artifact occurring due to the digital nature of the conversion. In a more extreme case, only odd or even digits may be displayed. Hence the resolution is specified separately from the display count.

The increased resolution of 20000 and 40000 count DMMs does not come without penalties. Longer settling times are required for the far right digits to reach their final value.

2.5.3 Sensitivity

This is the smallest change of the measured signal that can be detected. It depends on both resolution and the lowest measurement range of the instrument. For example, the sensitivity of a 5-1/2-digit DMM on the 200 mV range is 1 μV. (On a 200 mV display reading 200,000, the last zero is the one microvolt digit.)

2.5.4 Accuracy, Uncertainty and Repeatability

Accuracy is the largest allowable error that will occur under specific operating conditions. In other words, it is an indication of how close the DMM's displayed measurement is to the actual value of the signal being measured.

Accuracy for a DMM is usually expressed as a per cent of reading. An accuracy of one per cent of reading means that for a displayed reading of 100 volts, the actual value of the voltage could be anywhere between 99 volts and 101 volts. It may also be specified in parts per million (ppm).

Specifications may also include a range of digits added to the basic accuracy specification. This indicates how many counts the digit to the extreme right of the display may vary. So the preceding accuracy example might be stated as ±(1% + 2). Therefore, for a display reading of 100 volts, the actual voltage would be between 98.8 volts and 101.2 volts.

Accuracy is also specified over a period of time and within a temperature window (typically 1°C or 5°C around room temperature, with deratings for temperatures outside those ranges). Additional derating factors may apply for high voltage or current levels.

Accuracy for ac inputs is usually less than that for dc inputs. It is also optimized for 50-60 Hz. Other frequencies may have still poorer accuracy.

Analog meter specifications are determined by the error at full-scale, not at the displayed reading. Typical accuracy for an analog meter is ±2% or ±3% of full-scale. At one-tenth of full-scale, these become 20 per cent or 30 per cent of reading. DMMs with accuracy between ±(0.7% + 1) and ± (0.1% + 1) of reading or better, are available.

From a metrologist's standpoint, what appears in the manufacturer's specification sheet under accuracy is more properly deemed *uncertainty* with accuracy being reserved to indicate the probability of the reading being accurate. A specification of 1% uncertainty would have an accuracy of 99%. Common practice among instrument manufacturers is to use the term accuracy, and for calibration labs to use uncertainty with both representing the same thing in real life.

Repeatability or ***precision*** is often more important than absolute accuracy when making a series of measurements. We want the DMM to read the same value each time a measurement is made of exactly the same value. Repeatability is not usually specified directly but a limit is implied by the accuracy specification since it must always be within the specified accuracy. It may or may not be considerably better than the specified accuracy and can vary considerably from one DMM to another, even of the same manufacturer and model. Component aging, battery condition, temperature, and warm-up time may all affect repeatability to some degree. The only way to get a feel for a particular DMM's repeatability is to make a large number of measurements of a precision source under a variety of conditions. The source should have better than four times the rated accuracy of the DMM to be able to sort out repeatability and accuracy limitations.

Precision is given by the following equation:

$$\text{Precision} = \frac{1 - |X_n - \text{av}(X_n)|}{|\text{av}(X_n)|}$$

where X_n is the value of the *n*th measurement, av(*Xn*) is the average value of the set of *n* measurement.

For example, if we are monitoring a constant voltage of 1 V, and notice that the measured value changes by 20 µV between measurements, then the measurement precision is

$$\text{Precision} = \frac{1 - 20\,\text{mV}}{1\,\text{V}} \times 100 = 99.998\%$$

This specification is most valuable when we are using the voltmeter to calibrate a device or performing relative measurements.

2.5.5 Speed and Settling Time

Every meter has a settling time associated with its input circuit. The reading rates or measurement speeds of instruments are independent of the settling times. For high resolution meters, it may be necessary to allow time for input settling so as to achieve full-rated accuracy.

2.5.6 Average/True Root Mean Square (TRMS)

One of the most basic tasks of a DMM is measuring voltage. The waveforms associated with ac voltages are either sinusoidal, or non-sinusoidal (square, triangle, pulse, etc.). Quality DMMs display the rms value of these voltage waveforms. The rms value is the effective or equivalent dc value of the ac voltage. Most DMMs are "average responding," giving accurate rms readings if the ac voltage signal is a pure sine wave. They measure the average of the absolute value of ac voltage and are calibrated so that the readings are corrected to that of the rms value for a sine wave. This works well but errors occur if harmonics are present with the effect becoming progressively worse as the harmonic content increases. Table 2.4 shows the readings that would be obtained when average and TRMS responding DMMs are used to measure sine, triangular, square, pulse waveforms, and dc.

Table 2.4 Average vs. TRMS comparison of typical waveforms

Waveform	*Crest factor*[1]	*Form factor*[2]	*Actual pk-pk*	*Actual peak*	*Actual rms*	*TRMS* ac + dc	*TRMS* ac only	*Avg reading*	*Error for avg.*
dc	n/a	n/a	n/a	n/a	1.000	1.000	0	0	0%
sine	1.414	1.000	2.000	1.000	0.707	0.707	0.707	0.707	0%
triangle	1.732	1.155	2.000	1.000	0.577	0.577	0.577	0.555	–3.8%
square	1.000	1.000	2.000	1.000	1.000	1.000	1.000	1.111	+11.1%
sine +1 Vdc	2.829	1.731	2.000	2.000	1.224	1.224	0.707	0.707	–42.2%
pulse* (25%)	2.000	1.202	1.000	1.000	0.500	0.500	0.433	0.416	–3.8%
pulse* (12.5%)	2.833	1.453	1.000	1.000	0.353	0.353	0.331	0.243	–26.5%
pulse* (6.25%)	4.000	1.923	1.000	1.000	0.250	0.250	0.242	0.130	–46.2%

*Positive-going pulse between 0 and 1 V, [1]Ratio of peak to rms, [2]Ratio of rms to average

In fact, an rms measurement is a measure of the equivalent heating effect produced by a voltage and, to be accurate, must include any dc component present along with the ac component that most users associate with rms readings. Many TRMS DMMs are not capable of measuring the TRMS value of the combined ac + dc, and errors may result if there is any dc superimposed on the ac voltage. A few DMMs offer the ability to select average, ac TRMS, or ac + dc TRMS, thus allowing the user to choose the optimum mode for the desired measurement.

True rms conversion is normally performed by a dedicated chip. The conversion follows the basic definition of rms as shown in Figure 2.44.

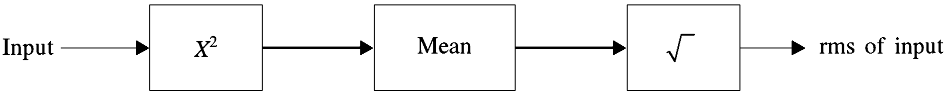


Figure 2.44 The rms conversion block diagram.

One implementation approach of this conversion is shown in Figure 2.45.

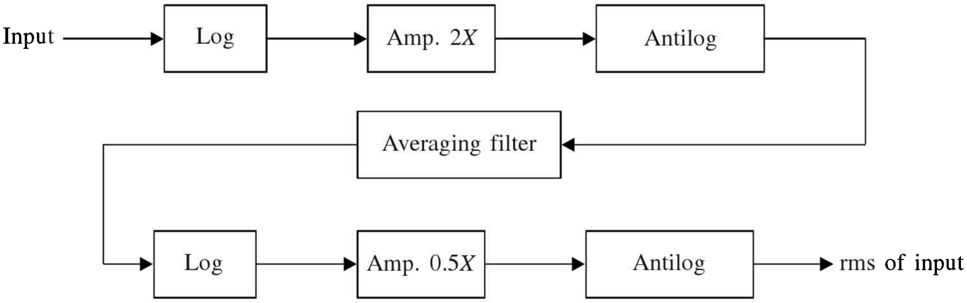


Figure 2.45 The rms conversion technique.

First the waveform is transformed into the log domain and then doubled. This effectively squares the input since doubling in the log domain is the equivalent of squaring in the linear domain. An antilog function transforms the waveform back into the linear domain and the mean value is derived using a low-pass filter. Finally, the waveform is transferred to the log domain and halved which is the equivalent to taking the square root in the linear domain. The final antilog function transforms the waveform to the linear domain resulting in the square root of the mean of the squared input.

2.5.7 Crest factor

Crest factor also should be considered when making ac voltage measurements of non-sinusoidal waveforms. It is defined as the ratio of the peak or *crest* voltage to the rms voltage. The crest factor relates to ideal waveforms as shown in Table 2.4.

The true rms reading DMMs usually specify the maximum crest factor that they can handle accurately. This usually be the lowest near the full-scale reading due to the saturation characteristics of the TRMS converter chip. Better accuracy may be obtained near the mid-scale.

This specification is important because it indicates the maximum peak value of an arbitrary waveform that the DMM can handle without overloading. The crest factor also affects the accuracy of the ac measurement. For example, given a certain DMM with an ac accuracy of 0.03% (this is always specified for sine waves), and as an additional error of 0.2% for crest factors between 1.414 and 5, the total accuracy for measuring a triangular wave (crest factor = 1.73) is 0.03% + 0.2% = 0.23%.

2.5.8 Form Factor

Form factor is defined as the ratio of the rms value to the average value and is therefore the factor which when multiplied by the average value of a waveform will equal the rms value.

2.5.9 Normal Mode Rejection Ratio (NMRR)

This ratio describes the ability of an instrument to reject a normal (differential) signal, and is given in dB by the equation

$$\mathrm{NMRR} = 20\log\left(\frac{V_{\mathrm{measured}}}{V_{\mathrm{in}}}\right)$$

where V_{in} is the voltage applied differentially to the instrument inputs, and V_{measured} is the value indicated by the DMM. This specification is useful for measurement systems that have filters to eliminate signals at a given frequency or over a range of frequencies. For systems that do not have filters, the NMRR is 0 dB. This specification often used to indicate the capability of the instrument to reject 50 or 60 Hz, is valid only at the specified frequency and useful only when making dc measurements.

For example, if we are measuring 1 mV dc with a DMM that specifies a NMRR of 130 dB at 50 Hz, and we have a normal-mode interference (noise) of 100 mV_{rms}, then the resulting measurement error is

$$\text{Measurement error} = 10^{(-130/20)} \times 100 \text{ mV} = 31.6 \text{ nV}$$

which is 0.003 per cent of the measured signal instead of the 10,000 per cent error that the 100 mV interference implies.

2.5.10 Common Mode Rejection Ratio (CMRR)

It is a measure of the capability of an instrument to reject a signal that is common to both input leads. For instance, if we are measuring a thermocouple voltage in a noisy environment, the noise from the environment appears on both the input leads. Therefore, this noise is a common-mode signal that is rejected by the CMRR of the instrument. The CMRR is defined by the equation

$$\text{CMRR} = 20 \log\left(\frac{\text{differential gain}}{\text{common mode gain}}\right)$$

This specification is very important, because it indicates how much of the common-mode signal will affect the measurement. CMRR is also frequency dependent. An equivalent equation to represent CMRR is

$$\text{CMRR} = 20 \log\left(\frac{V_{\text{in, measured}}}{V_C}\right)$$

where $V_{\text{in, measured}}$ is the indicated value for an applied common-mode voltage V_C.

2.5.11 Effective Common Mode Rejection Ratio (ECMRR)

This ratio is only valid for dc measurements and is the sum of CMRR and NMRR at a given frequency. It is the effective rejection on a given noise signal that is applied to both input leads because it is rejected first by the CMRR capability of the instrument and then again by its NMRR capability. This specification is mostly useful at the power line frequencies, where most of the noise resides.

2.5.12 Zero or Offset

All dc measurements are affected by residual offsets that may be in volts, amps or ohms. However, it is very unusual (in the context of DMM calibration) to find offsets of such large amplitude. Typically, offsets would be in mV, mA or mΩ. In a voltage measurement, placing a copper short across the voltage input terminals of the DMM results in a reading of exactly zero volts. However, thermally generated offsets, resulting from the use of dissimilar metals in the DMM's internal electronic circuits, result in a small net voltage offset that may amount to several tens of mV. Such offsets are normally relatively stable in a controlled environment and most DMMs have some form of offset compensation or input zero function that may be used to remove their effects. It is very important that the effects of offsets are removed from the measurement, otherwise they introduce a fixed error (in terms of volts, amps or ohms) throughout the measurement range.

2.5.13 Frequency Response

Traditionally, frequency response is interpreted as the –3 dB point, i.e. the high frequency point at which the reading has reduced to 70.7% of its nominal low frequency value. The manufacturer may employ various hardware and software adjustments to the response to make the –3 dB point as high as possible. This can result in undesirable peaks and troughs at lower frequencies. A more meaningful measure of response is *flatness* which is a measure of the deviation from an assumed flat response over specific frequency bands and may be specified in ppm or % rather than dBs. Depending on the rms conversion technique used, a DMM's ac response can suffer from low frequency problems as well as high frequency roll-off.

2.5.14 Input Resistance

This is an important parameter because it describes the potential loading effect of the DMM on the test circuit. For dc voltages up to 20 V, the input resistance may be greater than 10^{12} Ω and is the result of using feedback and FET devices. Note that some DMMs have a significant input capacitance on their dc voltage function. At higher voltages up to 1 kV, an attenuator is switched in to divide the applied voltage down to 10 V or 1 V. The attenuator typically has a resistance of 10 MΩ. For ac voltage, the resistance is usually lower—typically 1 MΩ in parallel with a 150 pF capacitor. Note that at high frequencies the input impedance reduces.

2.5.15 Input Bias Current

Amplifiers are usually designed for high gain, low noise, high bandwidth, high input impedance, low output impedance, and low input voltage offsets. However, all bipolar junction devices are essentially current operated. That is to say they require an input current to *bias* them into a desirable (usually linear) portion of their operating curve. Even field-effect devices may exhibit some internal leakage current back to their power supplies. Sensitive DMM designs seek to minimize these leakages by component selection and careful circuit design. However, there is inevitably some residual input current, which, unless preventative action is taken, flows in the external measurement circuit. A small current of 10 pA flowing in a 1 MΩ resistance develops a voltage of 10 mV. A modern sensitive DMM is expected to have an input bias current of < 50 pA.

2.5.16 Overload Protection

This is a measure of electrical ruggedness, and should be sufficient to protect the meter from line voltages commonly encountered. Typically, the ranges most susceptible to high voltage are the lowest voltage range (200 mV) and the ohms ranges. Similar to overload protection there is a maximum common mode voltage at which the meter can be used. This is the maximum voltage from earth ground that the input LO or COMMON terminal can withstand safely. The input LO terminal should always be at the lowest impedance point to minimize ground loops, and whenever possible, should be connected to circuit COMMON for safety.

2.5.17 Autoranging

DMMs provide autoranging facility wherein the input voltage is compared with the full-scale of the next lower range. If the present reading is less than the full-scale of the next lower range, the attenuation is reduced, otherwise it is increased. We must consider the autorange time when deriving system throughput because changing ranges increases the testing time.

2.5.18 Autozero

This is a selectable function. When this function is activated, the DMM measures internal voltages to maintain stability and accuracy as temperature changes over time. How often this is performed affects the reading rate. It is possible to disable this function, do an autozero every few hours, or program the DMM to autozero during the test fixture's unload/load cycle. Such test program changes can increase throughput.

2.5.19 Four-wire Remote Sensing

Normally to measure a resistance with a DMM, only two test leads are used as shown in Figure 2.46. DMM outputs (sources) constant current through the output terminals, which flows through the unknown resistance, R_U. The voltage across the same two terminals is then sensed (measured) by the meter and the reading displayed in ohms.

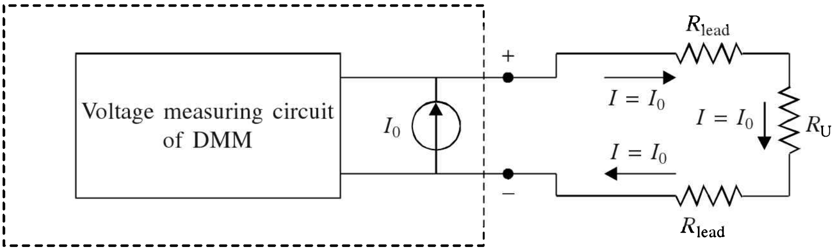


Figure 2.46 Two-wire resistance measurement.

When using only a single pair of leads, significant voltage drops can occur as a result of resistance in the feed cables, resistor interconnections, and terminations, meaning that the voltage delivered to the resistor at the end of the cable is not the expected value. Similarly, the measured signal is lower than the actual level at the resistor. DMM input resistance is normally very large. The measured voltage is given by the equation

$$V_{\text{measured}} = I(R_U + 2R_{\text{lead}})$$

Therefore,

$$R_{\text{measured}} = \frac{V_{\text{measured}}}{I} = R_U + 2R_{\text{lead}}$$

The error in measurement is $2R_{\text{lead}}$.

In a four-wire remote sensing, we use four test leads. One pair is used for the injected current, the other for sensing the voltage across the unknown resistor. Ideally, these measurements are made by a high-input impedance DMM that is tightly integrated with the source. The source then adjusts its output current automatically and transparently to ensure that the specified voltage or current is applied to the unknown resistor, regardless of cable lengths and interconnection losses. A higher accuracy is achieved because the resistances associated with the test leads are removed from the measurement path as indicated in Figure 2.47. This more accurate method is recommended for measuring resistances below 100 ohms.

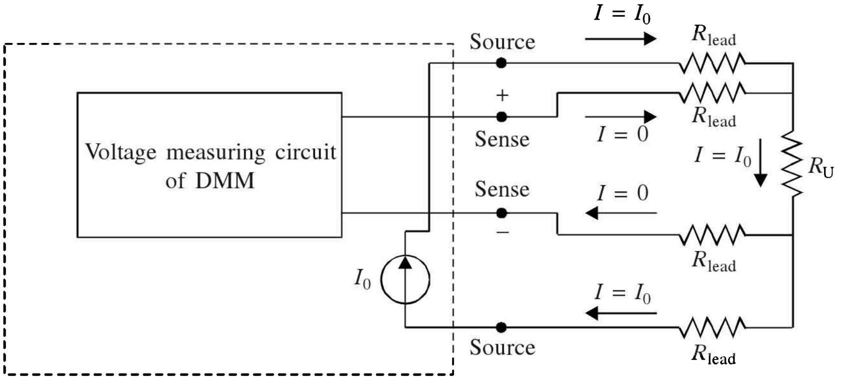


Figure 2.47 Four-wire remote sensing.

As shown in Figure 2.47, no current flows through the sense leads. Hence the DMM measures the voltage developed across the unknown resistor only. Therefore,

$$R_{\text{measured}} = R_U = \frac{V_{\text{measured}}}{I_0}$$

2.5.20 Environmental Specifications

The environmental concerns for DMMs are primarily temperature, humidity, altitude, shock and vibration. The operating temperature range is usually limited by the LCD, especially in the hand-held DMMs. At low temperatures, LCDs become sluggish and hard to read. At high temperatures, contrast may degrade due to the change in optimum LCD operating voltage with temperature, especially when multiplexed. Exceeding the non-operating temperature range may result in permanent deformation of the LCD polarizer, particularly when combined with high humidity.

Also, the specified accuracy is valid only for a relatively narrow temperature range of 23 ±5°C and for relative humidity less than 70 or 80%. The instrument is certainly usable over a much wider range but its accuracy will be degraded to some degree. Most DMMs are relatively well protected against humidity; however, they are not hermetically sealed and excessive humidity may cause internal condensation and possible loss of accuracy.

Altitude is not usually a problem, at least at normal measurement temperatures since icing is unlikely at these levels.

Shock and vibration are important for hand-held DMMs. Their small size and long test leads could result in their being accidentally dropped from bench top height. Consequently they are designed to survive a 3 to 4 feet drop onto a hard surface. These are normally in terms of a range of amplitudes and frequencies with related environmental test procedures.

2.5.21 Safety Ratings

Historically, each country has developed its own safety requirements for electrical devices sold within their country. They were not necessarily legal requirements but were increasingly dictated in recent years. The proliferation of differing national standard writing bodies (UL, CSA, EN, DIN, etc.) and international distribution of products led to the need for *harmonization* of the standards. This lead is now being provided by the International Electrotechnical Commission (IEC).

Much changing of standards has occurred during the 1990s. In the United States, UL 1244 is being replaced by ANSI/ISA S82.01-1994 and UL 3111-1, and in Canada, CSA C22.2, No. 231 is being replaced by CSA C22.2 No. 1010.1-92. All of these recent standards are harmonized with IEC 1010-1, which replaced IEC 348. IEC 1010-1 in turn is being adopted in the European Union as EN 61010-1. Note the similarity of the numbers for these standards. All contain 1010-1 in their designation with the exception of UL 3111-1. In this case, the 1010 designation had been used previously by Underwriters Laboratories (UL) for another standard and was unavailable. In any event, new DMMs must conform to an IEC 1010-1 harmonized standard. New instruments are now certified to these standards. The instruments manufactured earlier may not have been certified to IEC 348, UL 1244, CSA 231, etc. as it was not mandatory to do so at that time.

In the United States, the marking "NRTL" indicates a Nationally Recognized Testing Laboratory. These laboratories perform safety testing equivalent to UL standards.

Additionally, DMM inputs must now be marked with an overvoltage CAT (category) I, II, or III for protection against expected transients. These refer to the following usage:

- CAT I: Signal level, special equipment or parts of equipment, telecommunication, electronics
- CAT II: Local level mains, appliances, portable equipment
- CAT III: Distribution level mains, fixed installation
- CAT IV: Service drop to building (outside)

Similar overvoltage transients are expected for a 600 V CAT III or 1000 V CAT II mains line but a meter input designed for 600 V CAT III may be internally limited and not suitable for

1000 V CAT I or II use. It is therefore important to observe any voltage and current limitations indicated on the DMM or in its manual. Damage to the DMM or injury to the operator may occur if these ratings are exceeded. The current ranges are usually protected against overcurrent by fuses. Fuses of identical type must be used for ease of replacement. They are rated for the voltage indicated on the DMM or in its manual. Exceeding this voltage or replacement with a lower voltage fuse can result in a hazardous sustained arc if the fuse opens. Certification of products is not necessarily performed by the standards formulating organizations themselves. Other independent testing laboratories such as ETL are internationally recognized and perform their testing to UL, CSA, or IEC standards.

2.5.22 Electromagnetic Compatibility (EMC)

Instruments sold internationally now must also meet strict Electromagnetic Compatability (EMC) requirements, both for generation of emissions and for immunity to nearby electromagnetic and electrostatic fields. The EMC standards used for DMMs have evolved from local standards to different international standards and, finally, to the harmonized EN standards. DMMs as per European Union (EU), standards conform to EN 55011 for radiated emissions, EN 50082-1 for susceptibility, and display the CE conformity mark. EN 50082-1 is the generic electromagnetic immunity standard. It refers to a series of IEC standards for the individual test procedures. Those pertaining to DMMs are IEC 801-2-1984 for ESD immunity testing and IEC 801-3-1984 for RF immunity. Of these, the ESD immunity is probably of most interest to average DMM users since it assures that static zaps of this frequently handled instrument would not kill them. The normal testing is for protection to 8 kV. Testing to 15 kV is better as this is a realistic value that may be encountered on a dry day at an indoor location. IEC 1000-4-2 and IEC 1000-4-3 have now replaced EN 50082-1.

2.5.23 Additional Measurement Features

In addition to the basic voltage, current, and resistance measurement capabilities, modern DMMs may have many more functions. The common functions are autoranging, true rms ac voltage, frequency, capacitance, diode and continuity checking, max/min/avg reading, and temperature via both thermocouples and resistance temperature devices (RTDs). Less common functions are readings in decibels, displaying the difference between two channels, datalogging, and power and LCR bridge measurements. Some of these functions are explained below.

Capacitance

Many of the new features offered in DMMs are incorporated to satisfy specific requirements. For example, maintaining motor drives involves measuring voltage and current, but it also may be necessary to determine if a starting capacitor is faulty. We may need to check the temperature rise of the output devices under load. A DMM with both temperature and capacitance measuring functions allows us to perform a more complete test than is possible with a basic volts/amps/ohms DMM.

DMMs measure capacitance in accordance with the $Q = CV$ equation. A constant current charges the unknown capacitor for a known time, and the resulting voltage is measured. Alternatively, the charging continues to a set voltage, and the elapsed time is measured. Some DMMs approximate a constant current source by a dc voltage and a large resistor, especially on lower capacitance ranges.

Although we may be able to measure capacitance accurately enough to address a motor-drive application, a DMM would not be able to tell the capacitor's dissipation factor. Similarly, most hand-held DMMs do not measure inductance at all, so we cannot certainly determine the Q of an inductor with a DMM.

Temperature

Temperature measurement presents its own set of challenges. RTDs are often made from platinum wire and are relatively expensive. On the other hand, RTDs are linear, and measuring temperature with an RTD is equivalent to measuring resistance.

Thermocouples, though not expensive, require linearization and a stable temperature reference. The cold junction in a thermocouple circuit ideally is held at a constant temperature. Usually, one junction is kept in the environment in which the meter is to be used and the cold-junction circuitry is designed accordingly. A bench DMM may provide a better cold-junction compensation as well as a longer thermal time constant than are possible in a portable unit. Both design features support more stable and accurate readings.

A hand-held DMM should measure the cold-junction temperature so that the difference between it and the unknown junction temperature can be correctly indicated. A low-cost meter might be calibrated with the cold junction at 23°C, for example. Such a DMM would work well with large temperature differences in relatively stable ambient surroundings. It would be, however, useless for measuring small temperature differences, especially if we hold the meter in our hand which causes it to heat up.

Thermocouple use well-established polynomial equations to derive accurate temperature readings from millivolt thermocouple outputs. We need to make sure that DMM has the necessary processing power to solve the equations.

Some DMMs only measure a thermocouple's output voltage and leave the linearization to the user. DMMs that quote temperature accuracy use an integral microcontroller to perform the required mathematics. Some meters allow us to choose the type of thermocouple we wish to use. Others support only one type, such as the common K-type.

Frequency

Counting input pulses for a period of time provides a measure of frequency. Digital counters are used to count the pulses. Complications arise because input signals may not have perfectly square edges. This means that some amount of signal conditioning is needed between the DMM input and the actual digital counter.

The simplest approach is to limit or clip the input signal above a set level. This method works well with many sizes of signals. It fails when the input is too small and may give even false results when the signal is too large. Some clipping circuits fold back at high levels of overdrive, producing false signals equivalent to extra edges.

A more sophisticated design supports switching between ranges so that large signals can be attenuated or small ones amplified before the clipping stage. Additional cost and complexity accompany such higher-performance solutions. And greater complexity is required not just in the circuitry, but also in the user interface.

Diode

Some DMMs have a diode test mode. This mode measures and displays the actual voltage drop across a junction. For example, a silicon junction should have a voltage drop of less than 0.7 V when conducting in the forward direction and an open circuit in the reverse direction. This mode is therefore used to test and identify the anode and cathode terminals of the diode. Transistors can also be tested using this mode.

Interfacing ports

Modern DMMs provide IEEE 488 (GPIB) ports for interfacing it with other IEEE 488 instruments and computers. Some meters have RS 232 ports. Using these ports, various functions can be selected and measurements made remotely.

2.6 DMM CALIBRATION

2.6.1 DC Voltage Calibration Strategy

For a new DMM, the manufacturer tests every aspect of the instrument's performance. However, for routine repeat calibrations, the number of tests can be dramatically reduced if one accepts that instruments can be characterized according to their technology, operating principles and accumulated historical data. It is usually necessary to calibrate each dc voltage range in both polarities.

The *prime range*, where there is neither gain nor attenuation of the input voltage, is typically 10 V. This is the first range to be measured, as its errors will also be passed on to the other ranges. Calibration then progresses to the lower ranges of 1 V and 100 mV, then on to the 100 V and kV ranges. These are carried out last because of the effects of increasing power dissipation on both the DMM and the calibration source at the higher voltages. The order of the tests might be as follows:

Calibration sequence—dc voltage

10 V Zero, +10 V Gain, –10 V Gain
10 V Linearity ±1 V, ±5 V, ±15 V, ±19.9V
1 V Zero, +1 V Gain, –1 V Gain*
100 mV Zero, +100 mV Gain, –100 mV Gain*
100 V Zero, +100 V Gain, –100 V Gain*
1 kV Zero, +1 kV Gain, –1 kV Gain*

*Some DMMs may not have negative range adjustments.

Careful analysis of the results can reveal much about the instrument's performance. Variations in the error of 10 V range will also be passed on to other dependent ranges. Therefore, if the 10 V range increases by +5 ppm and the 100 V range shows no change, it means that the 100 V range has actually changed by –5 ppm, which was compensated by the 10 V range error. This kind of evaluation can be very useful for observing failures in specifications. If the 100 V range indicates a change of +10 ppm and the specification allowance is ±5 ppm, at first glance the 100 V range gain-defining components might be thought to be defective, when, in fact, the failure is caused by the instrument's internal reference. Either way, the 100 V range has failed against the specification, but the cause of the problem has nothing to do with the 100 V range components. Without this knowledge, the wrong components might get replaced, resulting in loss of history and additional expense.

2.6.2 Functional Tests

Essential functional tests are carried out before committing valuable time and effort to a potentially lengthy calibration process. Some microprocessor controlled DMMs have a diagnostic self-test function that is used to perform sophisticated internal checks. If such a facility is not available, each function and range is selected and a copper short, or a known voltage, resistance or current is applied to the DMM's input. This ensures that a stable reading is obtained and that the controls and displays are operating correctly.

Functional checks may be extended to cover input switching circuits, auxiliary outputs and remote-control digital interface, if applicable. If an IEEE 488 interface is fitted and the DMM is one of a type that should be regularly calibrated, a simple program is written to perform the functional testing. If a functional problem is found, a decision is made as to whether it is safe or not to proceed with a pre-adjustment calibration check. This is desirable if historical performance data is required, but it may also assist with fault diagnosis. However, if there is any risk of personal injury or damage to the instrument, the instrument should be repaired before proceeding with the calibration.

2.6.3 Input Characteristics

Periodically, a test is preformed to ensure that a DMM's input characteristics are correct. The test requires the use of an LCR bridge or relatively low-accuracy DMM. The object of the test is to ensure that there are no gross errors in the input characteristics. Note that it is essentially a functional test looking for percentage rather than ppm errors. For example, assume that a common ACV input resistance is 1 MΩ. The test then ensures that the input resistance is not excessively high or low—say, it is only within ±10%.

For most instruments, the ac input resistor forms a part of the input attenuator. Thus, significant changes in value are obvious when the calibration is verified. The DCV function input resistance measurement can be more difficult. For higher ranges, for example, 100 V to 1 kV, there is an input attenuator of typically 10 MΩ that determines the input resistance and voltage division ratio. Changes in the value of this component cause changes in the DCV calibration. However, for lower ranges, the input resistance may probably be determined by feedback and may be very high indeed. In fact, it may be so high that it may not be possible to measure reliably.

The usual reason for checking the input characteristics is to ensure that they do not significantly disturb the test circuit conditions. This is sometimes described as being non-intrusive. For current measurements, it is the *burden* voltage that is important. In a DMM, the burden is determined by the value of the current sensing resistors and can be measured as a resistance between the current sensing terminals.

Typically, the resistors are chosen such that the burden is 0.1 V or lower. For the resistance function, there are two parameters of interest. The first is the input resistance of the voltage sensing terminals (assuming 4-wire capability). This is usually very high indeed and may be the same as that of the lower DCV ranges. Any change in input resistance is directly reflected in the resistance function accuracy and, therefore, it is not normally necessary to measure it. The second parameter is the *compliance* of the DMM's current source. Ideally, the current flowing through the external resistance being measured should be independent of the voltage developed across the resistor—a true constant current source with infinite output impedance. However, if the current does vary, it affects the resistance function's linearity. This is discussed further in the resistance section.

2.6.4 Resistance Calibration Strategy

The resistance function of a DMM consists primarily of a constant current source providing a range of currents typically from 10 nA to 10 mA. Selecting a resistance range requires an appropriate constant current to pass through the unknown or standard resistance. The voltage developed across the resistance is then measured by the DMM's dc voltage function with an appropriate range set. The accuracy of the resistance function is dependent upon both the resistance current source (providing the resistance ranges) and the dc voltage ranges of the DMM.

Stability of the currents is typically controlled from the DMM's dc reference circuit. In this case, changes in the reference are common to both the current sourcing and voltage sensing, resulting in no change in resistance measurement accuracy. However, changes in the performance of either the current-defining resistors or voltage gain-defining resistors affect the resistance calibration. For this reason, the resistance function is usually measured after the dc voltage function.

For sensitive long-scale DMMs it is usually better to start with the high resistance ranges and work down. This allows a longer thermal stabilization time for the connecting leads and the DMM's internal circuits, which may be working at the 10 mA and 100 mV levels for the current source and voltage measurement circuits respectively. In any case, where a 4-wire capability is available, it should be used. Unless the DMM has a *true ohms* function, the effects of voltage offsets must be considered.

2.6.5 Resistance Function Linearity Tests

For DMMs with a resistance function accuracy of not better than 100 ppm (4-1/2 to 5-1/2 digit resolution), a decade resistor box could be used, but for the more accurate DMMs, resistance linearity can be more difficult to determine unless a selection of suitably accurate four-terminal resistances is available. One of the problems of trying to measure resistance linearity directly is the uncertainty of the individual resistor values. For example, to measure linearity on the 10 kΩ range of a 7-1/2 digit DMM with a maximum indication of 19.999999 kΩ requires several different resistance standards. Assuming that measurements are to be made at a minimum of five evenly spaced points throughout the range, for example, at 0, 5 kΩ, 10 kΩ, 15 kΩ and 19.9 kΩ, the difficulties in finding suitable standards soon become obvious.

Typically, the resistance standards are available at the normal decade values of 10 Ω (25 Ω may be available), 100 Ω, 1 kΩ, 10 kΩ, etc. and, therefore, do not provide even coverage throughout the range. When one considers that some DMMs have resistance linearity specifications of better than 0.5 ppm, and that individual resistance standards may have uncertainties of 1 ppm or more, test methods using separate resistors or decade boxes will be inadequate. For this reason, resistance linearity is not usually measured for routine calibrations of medium-to-high accuracy DMMs. For lower accuracy DMMs with up to 5-1/2 digits resolution, resistance calibrators suitable for resistance linearity testing are readily available. If it is necessary to determine the resistance linearity of a long-scale DMM, or confirm that the circuits are operating correctly, four-wire measurements can be used.

2.6.6 AC Voltage Calibration Strategy

The ac functions have the added dimension of frequency. This complicates calibration by introducing additional test points for each amplitude range. In a typical DMM, the ac measurements are made by an ac converter. This module must provide gain and attenuation for a wide range of signals, from typically a few millivolts to 1 kV. DMMs are available with 6-1/2 digit 100 mV ranges with a resolution of 100 nV. This resolution is only meaningful down to an input voltage level of perhaps 1 mV, below which, noise and linearity errors are likely to dominate the reading. It is likely that the limiting factor will be the noise of the calibration source rather than the DMM.

The gain of a DMM's ac function varies with frequency. This is known as its *frequency response* and requires that measurements be made at key points throughout each amplitude range. At low frequencies, from 40 Hz to 5 kHz, the DMM's gain is determined by resistor networks, its ac to dc converter and dc reference. At higher frequencies, reactive effects that are primarily capacitive, determine the *flatness* of the frequency response. Various means may be used to trim the response, for example, variable capacitance trimmers or software-controlled adjustment of the high frequency gain. The DMM manufacturer recommends certain adjustment points to compensate for LF gain and frequency response for each range. These adjustments should be *strictly* observed, as they are designed to give the best overall performance over the specified frequency range.

The prime range of the ac function is usually 1 V, with other ranges being obtained by gain or attenuation of the input signal. All the ac ranges depend on the rms converter.

Calibration sequence—ac voltage

1 V ac Zero, ±1 V dc Turnover*
Crest Factor (at 3:1 or 5:1 as required)*
1 V LF Gain, 1 V HF Gain, Check Frequency Response*

1 V LF Linearity, 1 V HF Linearity*

100 mV Zero, 100 mV LF, 100 mV HF, Check Frequency Response*

10 V Zero, 10 V LF, 10 V HF, Check Frequency Response*

100 V Zero, 100 V LF, 100 V HF, Check Frequency Response*

1 kV Zero, 1 kV LF, 1 kV HF, Check Frequency Response*

* These adjustments are usually iterative.

2.6.7 Current Calibration Strategy

A DMM measures current by sensing the dc or ac voltage developed across a current-sensing resistor or shunt. The same shunts are used for both dc and ac current. A separate pair of terminals is provided to simplify internal signal switching and minimize the impedance of the internal connections to the shunts. It is very much desirable to keep the burden voltage as low as possible, typically this is in the region of 100 mV.

The shunts are connected in series with suitable tapping points for voltage sensing. For example, to measure 0.1 mA, the current is passed through all of the shunts in series and their values are chosen such that the sum of their resistive values for any particular current range develops 100 mV for the specified current. This greatly simplifies the current switching, and also means that for any given current range, all of the shunts included will affect that range's performance. Knowledge of interdependence of the shunts can assist with drift and fault diagnosis. The voltage developed across the shunts is passed to the dc or ac 100 mV ranges, depending on whether the DCI or ACI function has been selected. Therefore, changes in the dc or ac 100 mV range performance also affect the current ranges.

Current calibration is normally the last function to be calibrated for the simple reason that knowledge of the DCV and ACV performance is essential before the DCI and ACI range performance can be evaluated. The order of calibration would normally be as shown below and is intended to minimize any heating effects by measuring the lowest current ranges first. Note that it is not a normal practice to switch alternately between DCI and ACI functions for each range, although this could be done. Note that some DMMs do not have an independent adjustment of the ac current ranges but rely on the dc gain adjustment, the *flatness* of the shunt's ac response and the accuracy of the ac voltage converter. In this case the ac voltage and dc current ranges must be adjusted first.

Calibration sequence—dc current

100 µ Zero, +100 µA, –100 µA

1 mA Zero, +1 mA, –1 mA

10 mA Zero, +10 mA, –10 mA

100 mA Zero, +100 mA*, –100 mA*

1 A Zero*, +1 A*, –1 A*

* These measurements may require a longer settling time due to self-heating and thermo-electric effects.

Calibration sequence—ac current

100 μA LF Gain, 100 μA HF Gain**
1 mA LF Gain, 1 mA HF Gain**
10 mA LF Gain, 10 mA HF Gain**
100 mA LF Gain, 100 mA HF Gain**
1 A LF Gain, 1 A LF Gain**

** Some DMMs do not have an HF adjustment.

2.6.8 Zero Considerations

All dc functions need their zero offsets evaluated and compensated before measurements of the gain values are made. Generally, it is the gain values that transfer traceability. This is because of the fact that zero is really only a baseline for the measurements. It is important that all dc measurements are referred to a known offset state, but that state does not have to be exactly zero. In fact, at normal room temperature it is very difficult to achieve a true zero in terms of volts, amps, or ohms. Consideration of zero offsets also include the calibration source, as all sources have residual offsets.

An offset affects all readings by a fixed amount. It is therefore very important that a "system" zero is performed when a DMM is being calibrated. A system zero means zeroing the DMM to the zero offset of the calibration standard. This removes the effects of the standard's offset from the measurement of the gain or full-range values. For voltage measurements the DMM should be zeroed to the source "zero" output.

For resistance measurements the DMM should be zeroed to the resistor such that the effects of voltage offsets in the High and Low circuits are removed. Current measurements require the DMM to be zeroed in the open-circuit state, although some calibration sources have a "zero" current output to which the DMM should be zeroed.

2.6.9 Generating a Test Plan

Routine calibration (no repair)

- Have any problems been reported?
- Functional checks, controls, etc.
- Measure prime dc and ac linearity
- Measure all gain values
- Adjust as required
- Check volt/hertz limits, etc.*

* Unless restricted use intended

Additional requirements after the instrument is repaired

- Adjust ADC if required
- Adjust line locking if required
- Adjust crest factor if applicable
- Check input characteristics
- Check common mode rejection ratio (CMRR)

After considering the requirements and the intended applications of the DMM, a test plan is devised. The plan should ensure that the basic calibration requirements of the instrument and those of the user are met with, that the measurements are made using standards of sufficient accuracy, and that the functionality and integrity of the instrument are verified. The plan also considers actions to be taken if the instrument needs to be repaired. It is likely that there will be two strategies depending on whether or not a fault condition is suspected or has been reported.

2.6.10 Measurement Uncertainty

The most significant contribution to the quality of DMM calibration is consideration of the measurement uncertainty. The sources of uncertainty can be the following:

Uncertainty contributions

Uncertainty of the standard. It is imperative that the calibration standard be of sufficient accuracy to be able to calibrate the DMM with confidence. Some manufacturers separate DMM specifications into two parts—the *calibration uncertainty* and the instrument's *relative accuracy*. If the available calibration uncertainty is larger (worse) than the specified calibration accuracy, the DMM's total accuracy specification will no longer apply. Sometimes test accuracy ratios (TAR) are quoted as a requirement, i.e. a TAR of 4:1 requires the accuracy of the standard to be 4 times better than the specification of the DMM. This is to ensure that the residual errors of the standard do not significantly affect the calibration accuracy of the DMM. Ideally, TARs could be applied to all measurements—including those made by national standards laboratories. However, it is not practical at this level or even necessary, if all sources of error have been identified and corrected and that a sound uncertainty analysis has been performed. Where such corrections are not applied, or where there is no calibration accuracy requirement specified, TARs are appropriate. For more details refer to Chapter 13 on "Standards in Quality Management".

Resolution of the DMM. The *resolution* or *scale-length* of the DMM determines the smallest change in the reading that may be observed. Clearly, it could become a limiting factor in the measurement, regardless of how accurate the calibration standard is. Assuming the calibration of 1 V dc, a 6-1/2 digit DMM with a scale-length of 1.000000 V can resolve 1 μV or 1 ppm of its nominal range, while a 4-1/2 digit DMM can only resolve 100 μV or 0.1% for the same conditions.

Short-term stability of the DMM with time and temperature. The short-term stability of the DMM (and the calibration standard) with time and temperature also affects the uncertainty of measurement. Usually the dominant factor here is temperature coefficient of the DMM and stability of the calibration environment. Drift is not usually significant unless the calibration takes several hours or the instrument is defective. An allowance for temperature coefficient and short-term stability can usually be obtained from the manufacturer's specification data.

Combined noise of standard and DMM. A more dominant uncertainty contribution is usually the noise or run-around of the DMM reading during the measurement. While it may be interesting to consider the individual noise contributions of the standard and the DMM, in practice it is their combined effect that is important. If individual readings can be easily observed, and if the noise is predominantly random, the sample standard deviation of the readings can be calculated and used as the uncertainty contribution for combined noise. It is important that the configuration of the DMM be representative of normal use for this measurement.

The calibration procedure. Finally, the calibration procedure itself has an influence on the measurement uncertainty. A poorly chosen test sequence, insufficient settling time or poor interconnection techniques—all introduce additional errors that may pass unnoticed by the operator. This is the reason why the manufacturer's recommended procedure is used as the basis for DMM calibration. Note that this also applies to the use of the calibration standards.

2.7 SELECTING A DMM

When selecting a DMM, accuracy, resolution, and sensitivity are important parameters to consider. At the same time, however, we should also look at other factors like speed, convenience, productivity enhancements, noise filters and other features to help overcome external noise problems, measurement channels, etc. Taking a broader view, we must analyze two areas—the device under test (DUT) and the test environment. By asking some questions as explained in Figure 2.48, we can easily assess measurement functions and performance requirements.

With each question, we evaluate how the DMM fits into our test environment. Three more questions to be added are:

- Does the DMM meet or exceed the requirements for our application?
- Will the DMM easily integrate with our test setup?
- Does it have a feature that no other DMM has that will cut our test time?

The answers are sometimes obvious in an instrument's data sheet, but they require a careful reading of the specifications.

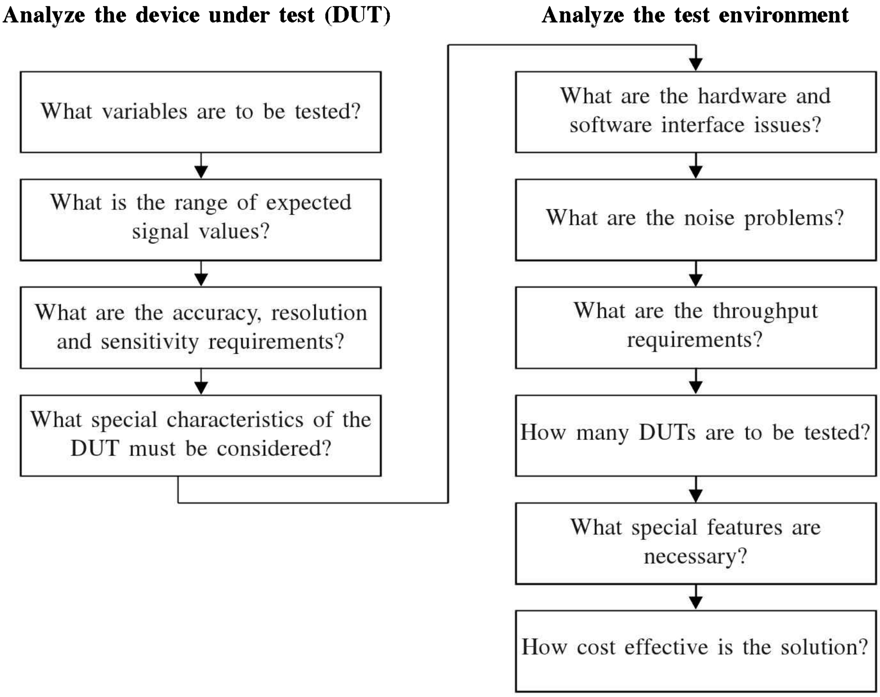

Analyze the device under test (DUT)
Analyze the test environment
What variables are to be tested?
What is the range of expected signal values?
What are the accuracy, resolution and sensitivity requirements?
What special characteristics of the DUT must be considered?
What are the hardware and software interface issues?
What are the noise problems?
What are the throughput requirements?
How many DUTs are to be tested?
What special features are necessary?
How cost effective is the solution?

Figure 2.48 Steps in selecting a DMM.

What variables are to be tested?

The first step is to analyze the DUT and the required measurements—what are the signal characteristics, and what data need to be collected? A typical five-function DMM measures dc and ac volts, dc and ac current, and two-wire ohms. More advanced DMMs may offer additional measurements such as:

- True rms measurements
- Four-wire ohms testing
- Temperature readings with a thermocouple or RTD
- dB and dBm
- Frequency and period
- Audio signal distortion
- Capacitance
- Checking diodes

These multifunction DMMs can reduce overall test equipment costs.

What is the range of expected signal values?

The DMM must have a wider measurement range than our expected minimum and maximum test signals. At the low end, we should pay attention to the dc noise specification. If the noise level is higher than the signal to be measured, we have to slow down measurements and use noise-reduction techniques, or choose a DMM with a lower measurement range and noise. At the high end, the maximum allowable input might include specifications for a volt-hertz product limitation or a time limit on ac signals. We make sure that the restrictions on ac measurements are acceptable.

Some DMMs have an over-range specification stated as a percentage of range, typically 120% on all but the top range. If the extra 20% lets us use a fixed range instead of autoranging, we can make our measurements faster.

For floating measurements, consider the DMM's common-mode voltage specification, the maximum voltage between the low input terminal and ground. For example, with 500 V between the high and low input terminals, and another 500 V between the low input and ground, the common-mode voltage specification must be at least 1000 V.

How does DMM handle accuracy resolution, and sensitivity?

A DMM's accuracy, resolution, and sensitivity define the unit's performance more than any other specification. DMM manufacturers typically have a family of instruments, with increasing order of performance in these areas along with the DMM's cost.

Accuracy is the degree of conformity of a measurement to a standard or true value. Manufacturers usually specify DMM accuracy as a per cent of a reading plus a per cent of the range (or a number of counts of the least significant digit). They may also specify it in ppm, where 1 ppm is the equivalent of 0.0001%.

Accuracy is also specified over a period of time and within a temperature window (typically 1°C or 5°C around room temperature, with deratings for temperatures outside those ranges). AC performance is typically specified over a range of frequencies. Additional derating factors may apply for high voltage or current levels. We should study the specifications to see if these factors must be considered at the signal levels of our test.

Resolution is the smallest portion of a signal that can be measured or displayed. We should study the specifications to see if these match our requirements.

Sensitivity is the smallest change of the measured signal that can be detected. It depends on both resolution and the lowest measurement range of the instrument. We should study the specificatons to see if the sensitivity meets our requirements.

What special characteristics of the DUT must be considered?

Review the DMM's specifications to make sure that it has functions that support all our requirements. For example, is our DUT susceptible to self-heating from the test current on ohms ranges? If so, the ability to control test current on each range is important. Also, for contact resistance measurements, dry circuit testing may be necessary, which limits test voltage, to prevent disturbance of oxidation on the contacts.

If DUT is a high-impedance device, a DMM may not be the best instrument to use because it will load the DUT.

What are the hardware and software interface issues?

The DMM must work with the line voltage and frequency used in our tests. But if we standardize on a model that can also be used in other countries, it should support a variety of power standards. A universal power supply that automatically senses voltage and frequency avoids problems that can arise if a DMM power switch must be manually set. We may also want a battery option for portable applications. Finally, check the power consumption (VA rating); we may need additional cooling in equipment racks.

Next, is this a stand-alone application, or does the DMM need data communications ports (e.g. RS 232 and IEEE 488)? Is the DMM going to replace older test equipment? If so, it may have to fit into existing rack space.

On software issues, what programming languages or drivers are supported? Will a new command set have to be learned? Some DMMs can emulate the command sets of other multimeters. This might eliminate the need to revise test programs and make drop-in replacements that much easier.

How will the DUT be connected to the test system? Is there a test fixture? DMMs typically have banana jacks and a variety of test lead terminators, such as alligator clips and spade lugs. Are those sufficient, or will adapters be necessary? Are safety test leads and special probes available for high-voltage signals?

What are the noise sources?

In industrial environments, instruments must contend with power line noise and random noise, such as fluorescent lights and microprocessor clocks. Within a switching system, relay contact bounce and high-voltage arcing may also be present.

How susceptible is the DMM to these noise sources? A DMM's specifications for normal-mode rejection ratio (NMRR) and common-mode rejection ratio (CMRR) indicate its noise susceptibility. Configuring the DMM's signal integration period to be an integer multiple of power line cycles reduces the power line noise. Digital filtering helps average out random noise from the measurements.

How much noise does the DMM emanate? This is especially critical when measuring low signal levels. We should look for compliance with common EMI standards.

What are the throughput requirements?

Measurement speed is important in many test situations, especially production environments, because it is a major determinant of throughput. Speed is usually stated as readings per second for a set of operating conditions. These conditions may include:

Measurement resolution. A 4-1/2-digit reading is acquired faster than a 7-1/2-digit reading. We should use only as much resolution as is needed. Some DMMs have programmable resolutions, which allows trade-offs between speed and accuracy.

Integration period. This is the window of time a signal is sampled by the A/D converter. It is often expressed in number of power line cycles (NPLCs), where 1 PLC for 60 Hz is 16.67 ms and 1 PLC for 50 Hz is 20 ms. NPLCs that are integer multiples (1, 5, 10, and so on) can reduce the most common type of normal-mode noise—50/60 Hz line pickup. The larger the value of *N*, the greater the line noise reduction; but the measurement takes a longer period of time.

Digital filtering. This is the quantity of A/D conversions averaged for each reading. It could be up to 100 conversions. Filtering stabilizes noisy readings but slows down measurements. The flexibility of programmable filtering can help reach a compromise between speed and accuracy.

Autozeroing. When this function is activated, the DMM measures the internal voltages to maintain stability and accuracy as the temperature changes over time. How often is this performed affects the reading rate. It may be possible to disable this function, do an autozero every few hours, or program the DMM to autozero during the test fixture's unload/load cycle. Such test program changes can increase throughput.

Besides reading rates, we must consider other speed specifications, such as the range changing rate and the autorange time, when deriving system throughput. Because changing ranges increases testing time, it is advantageous to stay on a fixed range. Check the specifications to see if the resolution and accuracy of a single range are sufficient. For multiple tests on a DUT, the function changing rate might be a throughput concern.

Also, what is the trigger latency? Hardware triggers are typically faster than software triggers. Some DMMs have a special microprocessor just to handle triggers. This shortens the trigger latency considerably.

Settling time (or response time) is another speed specification to be considered. For a measuring instrument, settling time is the time between an input signal change and indication of that change at a rated accuracy. Settling time is more of an issue when testing high-impedance devices.

How many DUTs are to be tested?

A device may have several test points, or there may be several devices to test. In either case, using relays to switch the signals eliminates moving test leads. Switching is particularly useful in a production test environment in which hundreds or thousands of devices are tested each work shift.

Switching can take place inside the DMM if it has an optional relay card, which typically expands the number of measurement channels to 10 or 20. Otherwise, an external switching system mainframe may be necessary.

Internal switching is usually limited to general-purpose signals, but external switching can handle most types and levels of signals that a DMM can measure. In any case, the relays or mainframe will create some signal degradation and additional settling time delays.

Regardless of where the switch card resides, these factors affect low-level signals unless we choose the type of relay card and system connections carefully. One advantage of DMMs with internal switch card slots is its simplified, less costly system integration.

What special features are necessary?

To differentiate product offerings, manufacturers constantly add new functions and features to DMMs. This benefits production test systems, where rack space is at a premium. Some features to look for include:

(a) Faster and simpler hardware triggering. Such improved triggering will reduce test time.

(b) An internal buffer. It is faster to let a DMM store data and send the readings all at once than send single readings to the computer.

(c) Stored setups for function and range changing. These quickly set up a DMM for different tests or devices and speed testing by minimizing external data bus traffic between an IEEE 488 instrument and the PC controller.

(d) Measurement limit testing and digital I/O for binning of DUTs. For example, a DMM with two sets of limits and a strobe signal can sort resistors into tolerance bands without intervention of the IEEE 488 controller.

(e) Functions for telecommunication testing. For example, we might need features for measuring total harmonic distortion in audio circuits.

How cost effective is the solution?

In addition to the purchase price, there are other factors that contribute to the cost of ownership. We should consider the startup costs including programming, system integration, and training. Are the software libraries and drivers included with the instrument, sufficient to get it up and running? On the other side, there might be an opportunity cost based on the throughput of a faster DMM.

There are also long-term costs for maintenance. These are associated with calibration intervals and necessary calibration equipment. For instance, will the DMM be sent to the manufacturer for calibration? Will it be necessary to purchase spare units to reduce production line downtime? How long is the warranty? All these factors affect cost of ownership.

One other step

After completing this stepwise process and narrowing the DMM choices to two or three models, we should discuss the measurement needs with the manufacturer of each shortlisted model to verify our selection or otherwise.

Bench-top DMMs with basic one-year accuracy ranging from 0.0006% to 0.03%, resolution from 4-1/2 digits to 8-1/2 digits, and sensitivity from 1 nV to 10 V are now available. By thoroughly evaluating our needs and all the product features, we can make an informed and cost-effective purchasing decision following these easy steps.

QUESTIONS

2.1 Draw the block diagram of a digital multimeter. Why is the crest factor important for ac measurements?

2.2 Explain the operation of the weighted resistor D/A converter and R-2R ladder D/A converter. Compare their performance.

2.3 (a) Explain the operation of a digital ramp A/D converter.

(b) An 8-bit digital ramp A/D converter has the following values:

Clock frequency = 100 kHz, Threshold voltage = 100 μV, $V_r = 2.56$ V

Determine the digital equivalent representation for a dc input of 1.728 V and also find the conversion time required for this converter.

2.4 (a) Explain the operation of the dual slope ADC.

(b) What are the advantages and disadvantages of using dual slope ADC in digital voltmeters?

2.5 A digital voltmeter uses a dual slope A/D converter. The reference voltage is 1 V, the clock frequency is 1 kHz and the counter is a 3-decade counter. If the integrating capacitor charges to 5 V in 1 s and discharges at the rate of 10 V/s, what is the input voltage if the counter has stored 1500 pulses?

2.6 (a) Explain the process of successive approximation used in a 3-bit ADC.

(b) Explain aliasing and the process of eliminating the same.

2.7 (a) Why is a sample and hold circuit used in ADC? Define the response parameters of a sample and hold circuit.

(b) The maximum charging current of the input opamp in a sample and hold circuit is 100 mA. If the expected droop rate is 1 mV/s and the acquisition time for 3 V step is 6 μs, calculate the droop current.

2.8 The speed of an 8-bit ADC is limited by the counter used in the converter. The converter has a maximum speed of 4×10^7 counts per second. Estimate the maximum number of A/D conversions per second that can be achieved. Will the process be actually faster or slower and why?

2.9 Explain the type of errors which affect the accuracy of an A/D converter and D/A converter.

2.10 Explain the operation of a sigma delta A/D converter. What are its special characteristics?

2.11 Show that the required number of bits, N, for the given reference voltage V_r, fixed in-band noise power, n_0^2, and the oversampling ratio, OSR, are given by the equation

$$N = -\frac{1}{2}\log\,(\mathrm{OSR})_2 - \frac{1}{2}\log\,(12)_2 - \frac{1}{2}\log\,(n_0^2)_2 + \log\,(V_r)_2$$

2.12 Explain the four-wire remote sensing used in digital multimeters.

Chapter 3

Analog and Digital Oscilloscopes

3.1 INTRODUCTION

The oscilloscope is basically a graph-displaying device. It draws a graph of an electrical signal. In most applications the graph shows how signals change over time: the vertical (*Y*) axis represents voltage and the horizontal (*X*) axis represents time. The intensity or brightness of the display is sometimes called the *Z*-axis (see Figure 3.1.) This simple graph can tell us many things about a signal. Here are a few:

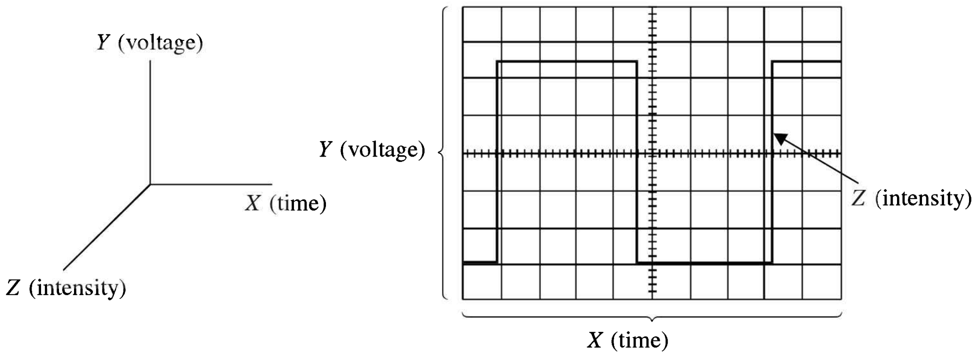


Figure 3.1 *X*, *Y*, and *Z* components of a displayed waveform.

- We can determine the time and voltage values of a signal.
- We can calculate the frequency of an oscillating signal.

- We can tell how often a particular portion of the signal occurs in relation to other portions.
- We can tell if a malfunctioning component is distorting the signal.
- We can find out how much of a signal is direct current (dc) and how much is alternating current (ac).
- We can tell how much of a signal is noise and whether the noise is changing with time.

An oscilloscope's front panel includes a display screen, and the knobs, buttons, switches, and indicators that are used to control signal acquisition and display. Front panel controls normally are divided into Vertical, Horizontal, and Trigger sections, and of course there are display controls and input connectors.

Oscilloscopes are used by everyone from television repair technicians to physicists. They are indispensable for anyone designing or repairing electronic equipment. The usefulness of an oscilloscope is not limited to the world of electronics alone. With the proper transducer, an oscilloscope can measure all kinds of phenomena. A transducer is a device that creates an electrical signal in response to physical stimuli such as sound, mechanical stress, pressure, light, or heat. An automotive engineer uses an oscilloscope to measure engine vibrations. A medical researcher uses an oscilloscope to measure brain waves. Let us enumerate the types of oscilloscopes.

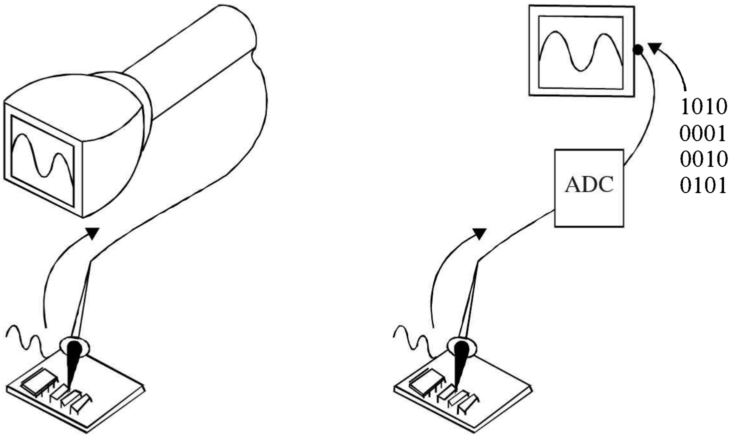
3.2 TYPES OF OSCILLOSCOPES

Oscilloscopes come in analog and digitizing types. Fundamentally an analog oscilloscope works by applying the measured signal voltage directly to an electron beam moving across the oscilloscope screen (usually a cathode-ray tube, CRT). The back side of the screen is treated with a phosphor coating that glows wherever the electron beam hits it. The signal voltage deflects the beam up and down proportionally, tracing the waveform on the screen. The more frequently the beam hits a particular screen location, the more brightly it glows. This gives an immediate picture of the waveform.

The range of frequencies an analog scope can display is limited by the CRT. At very low frequencies, the signal appears as a bright, slow-moving dot that is difficult to distinguish as a waveform. At high frequencies, the CRT's writing speed defines the limit. When the signal frequency exceeds the CRT's writing speed, the display becomes too dim to see. The fastest analog scopes can display frequencies up to about 1 GHz.

In contrast, a digitizing oscilloscope uses an analog-to-digital converter (ADC) to convert the voltage being measured into digital information (Figure 3.2). The digitizing scope acquires the waveform as a series of samples. It stores these samples until it accumulates enough samples to describe a waveform, and then re-assembles the waveform for viewing on the screen. The conventional digitizing scope is known as a DSO (Digitizing Storage Oscilloscope). It uses raster screen.

Recently a third type of oscilloscope architecture called Digital Phosphor Oscilloscope (DPO) has emerged. The DPO is a digitizing scope that faithfully emulates the best display



An analog oscilloscope traces signals

A digital oscilloscope samples signals and constructs displays

Figure 3.2 Analog and digitizing oscilloscopes.

attributes of the analog scope and provides the benefits of digital acquisition and processing as well. Like the DSO, the DPO uses a raster screen. Additionally, the DPO also provides an intensity-graded trace.

For both DSOs and DPOs, the digital approach means that the scope can display any frequency within its range with equal stability, brightness, and clarity. The digitizing oscilloscope's frequency range is determined by its sample rate, assuming that its probes and vertical sections are adequate for the task.

For many applications either an analog or a digitizing oscilloscope can be used. However, each type has unique characteristics that may make it more or less suitable for specific tasks.

Analog oscilloscopes are preferred when it is important to display rapidly varying signals in real time (as they occur). The analog scope's chemical phosphor-based display has a characteristic known as intensity grading which makes the trace brighter wherever the signal features occur most often. This makes it easy to distinguish signal details just by looking at the intensity levels of the trace.

Digitizing storage oscilloscopes allow to capture and view events that may happen only once, for example, transients. Because the waveform information is in digital form, it can be analyzed, archived, printed, and otherwise processed, within the scope itself or by an external computer. The waveform does not need to be continuous; even when the signal disappears, it can be displayed. However, DSOs have no real-time intensity grading, therefore they cannot express varying levels of intensity in the live signal.

The Digital Phosphor Oscilloscope breaks down the barrier between the analog and the digitizing scope technologies. It is equally suitable for viewing high frequencies or low repetitive waveforms, transients, and signal variations in real time. Among the digitizing scopes, only the DPO provides the Z (intensity) axis which is not available in conventional DSOs.

Although analog oscilloscopes work somewhat differently than digitizing oscilloscopes, some of the internal systems are similar. Analog oscilloscopes are simpler in concept and are described first, followed by a description of digitizing oscilloscopes.

3.3 ANALOG OSCILLOSCOPES

When we connect an oscilloscope probe to a circuit, the voltage signal travels through the probe to the vertical system of the oscilloscope. Figure 3.3 is a simple block diagram that shows how an analog oscilloscope displays a measured signal.

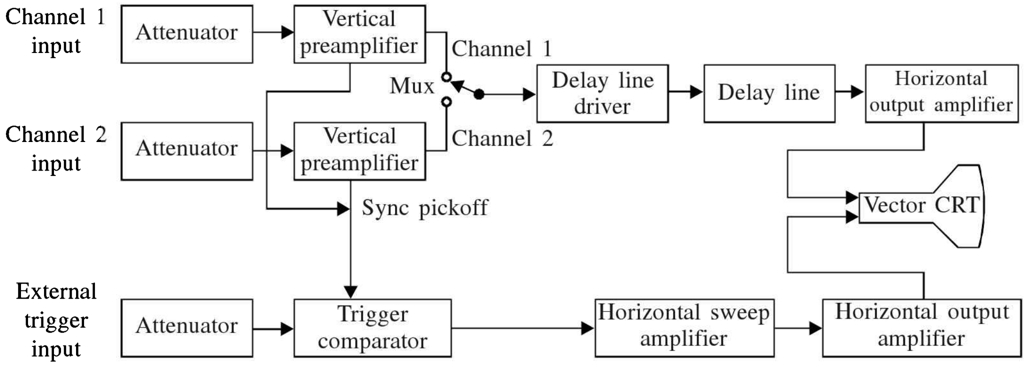


Figure 3.3 Block diagram of on analog oscilloscope.

As we can see in Figure 3.3, an analog scope has two major signal paths. The first is the vertical signal path, which ultimately is responsible for deflecting the CRT beam vertically in response to the input signal. The second path is the horizontal. It triggers the scope and moves the beam from left to right across the screen. In a typical display, time is represented horizontally and voltage is represented on the vertical axis.

The signal is first attenuated, if required, using the attenuators. Attenuators match the high impedance of the scope probes (typically 1 MΩ or 10 MΩ) to the low impedance of the vertical preamplifiers. The attenuators also scale the input signals to a level the vertical preamps can handle. The amount of attenuation and preamp gain is set by the vertical sensitivity knob provided on the front panel.

The triggering portion plays a very important part in the operation of a scope. It determines where (in time) the trace starts. In essence, the triggering circuits tell the horizontal section when to start moving the beam from the left side of the CRT to the right. If the trace starts too early, a part of the signal would not be seen. The same is true if it starts too late. Figure 3.4 illustrates this.

How does the trigger circuit know when to trigger? It gets a replica of the signal called the *sync pickoff* from the selected trigger source. This sync pickoff is compared to a pre-set trigger voltage that is set with the front panel trigger level knob. Most analog scopes specify a slope as well as a trigger voltage. This allows the trigger to take place at a specific point on a rising or falling transition.

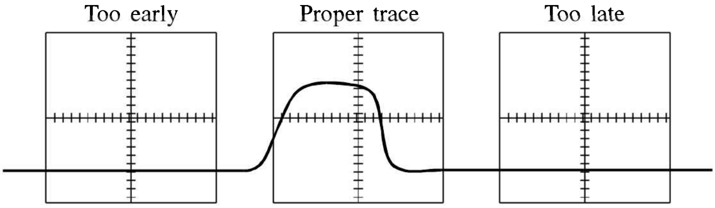


Figure 3.4 Triggering action.

When the trigger circuit finds a voltage and transition from the source that matches those set with the trigger controls, the beam starts moving from left to right. The speed of the beam is determined by the seconds/division knob on the front panel (also called the time base). As the beam is moved horizontally across the screen, the vertical amplifiers move the beam up and down, relative to the input voltage.

Both the horizontal sweep and vertical deflection information should arrive at the CRT at the same time. If they do not, the scope cannot display the voltage information properly. Since the delays in the horizontal path are longer, vertical information reaches the CRT before the horizontal information. The solution to the problem is to put a calibrated delay into the vertical path so that both horizontal and vertical signals reach the CRT at the same time. When properly adjusted, an analog scope does not display the signal fluctuations before the trigger event. Figure 3.5 shows the effect of a delay line.

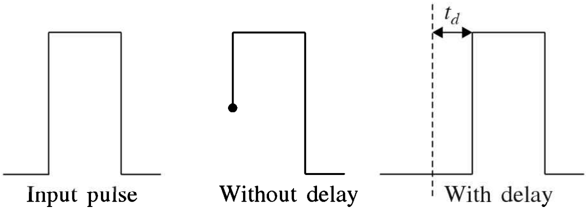


Figure 3.5 Delay line effect.

For external triggering, the horizontal path of an analog scope has an attenuator like the vertical channels. This attenuator serves the same purpose as that in the vertical channels, i.e. impedance matching and scaling the external trigger signal. However, the horizontal attenuator is followed by trigger comparison circuits, instead of a preamp as in the vertical channels.

The horizontal portion of the scope, which is responsible for moving the trace along the time or horizontal axis, directly affects the time accuracy of an analog scope. The horizontal beam movement is controlled by a voltage ramp (called the sweep); the time interval accuracy of the scope depends on this ramp.

Once the trigger comparator has found a valid trigger, it starts the horizontal sweep ramp generator. As the ramp rises, it causes the beam to move from left to right across the CRT. Since the left-to-right movement represents time on the CRT, the ramp must be truly linear. If the ramp has nonlinearities, the beam moves at different rates across the screen. This is shown in Figure 3.6. Typically, ramp linearity controls the time interval accuracy of an analog scope within ±3%.

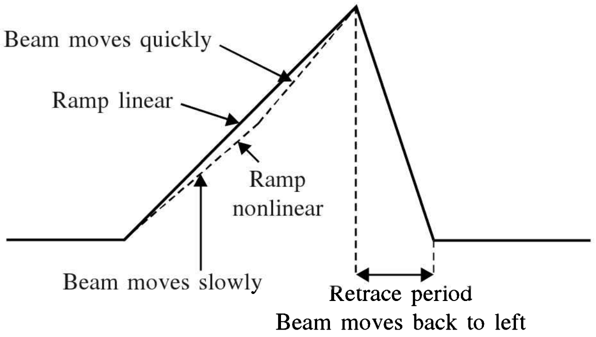


Figure 3.6 Effect of ramp nonlinearity.

The horizontal sweeping action and the vertical deflection action together trace a graph of the signal on the screen as shown in Figure 3.7. The trigger is necessary to stabilize a repeating signal. It ensures that the sweep begins at the same point of a repeating signal, resulting in a clear picture as shown in Figure 3.8.

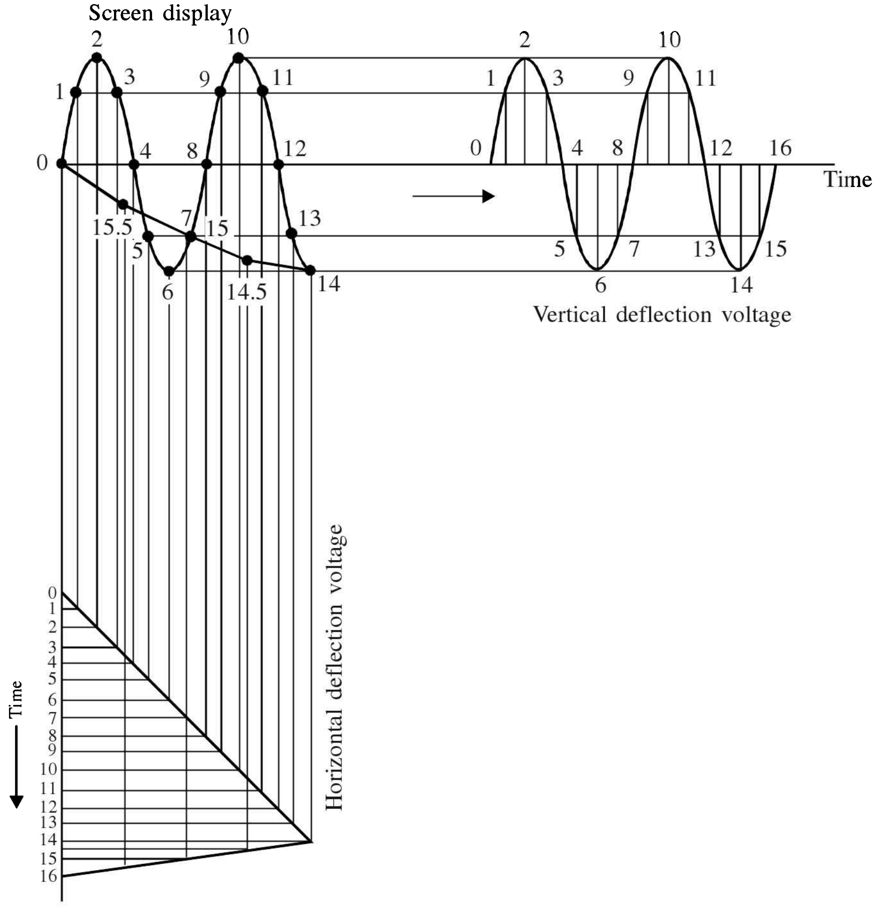


Figure 3.7 Display of the signal during one pulse of the sweep.

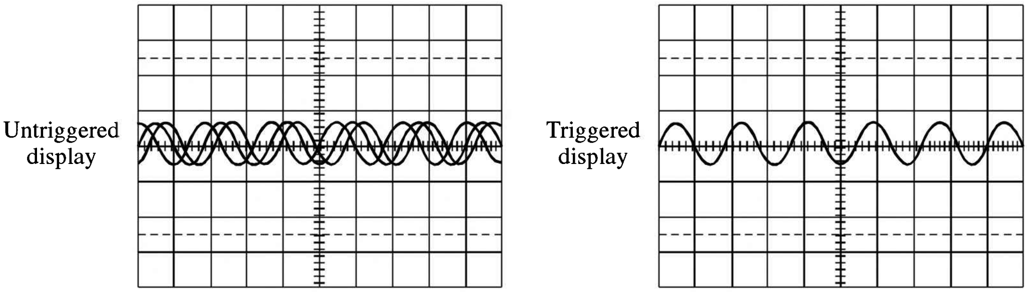


Figure 3.8 Triggering stabilizes a repeating waveform.

In brief, when using an analog oscilloscope (or any other type of oscilloscope), we need to adjust three basic settings to accommodate an incoming signal:

- *The attenuation or amplification of the signal.* We use the volts/div control to adjust the amplitude of the signal to the desired measurement range.
- *The time base.* We use the sec/div control to set the amount of time per division represented horizontally across the screen.
- *The triggering of the oscilloscope.* We use the trigger level to stabilize a repeating signal, or for triggering on a single event.

Cathode ray tube (CRT)

All CRTs have three main elements—an electron gun, a deflection system, and a screen. The electron gun provides an electron beam, which is a highly concentrated stream of electrons. The deflection system positions the electron beam on the screen, and the screen displays a small spot of light at the point where the electron beam strikes it. Figure 3.9 shows the details of a CRT. The components of the CRT as numbered in Figure 3.9 are described below:

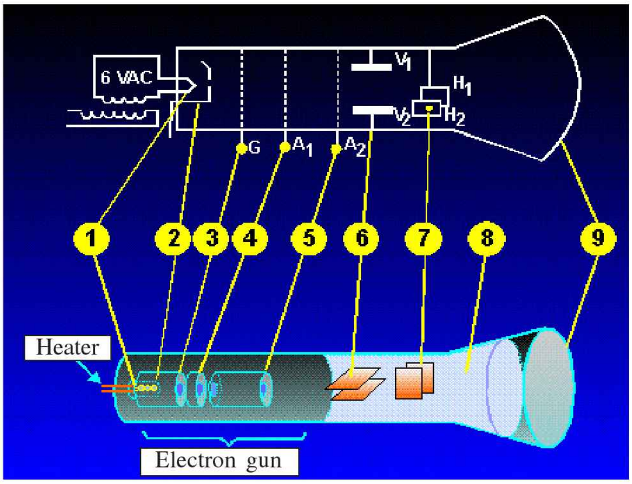


Figure 3.9 Details of a CRT.

Heater (1). It serves as the source of heat for the cathode in the CRT.

Cathode (2). It serves as the source of thermionically emitted electrons in the CRT. The cathode is circular in shape and the outer surface is coated to ensure that electron emission is roughly unidirectional.

Control grid (3). The control grid controls the number of electrons that will be fired across the *gap*. The CRT uses a cap-like grid. Intensity is controlled by controlling the grid voltage.

Focusing anode (4). It serves the dual purpose of attracting electrons from the area of the control grid and focusing the electrons into a beam. This is achieved by applying positive potential with reference to the cathode.

Accelerating anode (5). This anode is used to accelerate the electrons towards the front of the tube by the application of a high positive voltage.

Vertical-deflection plates (6). These plates move the electron beam up and down the screen. The input signal is applied to these plates.

Horizontal-deflection plates (7). These plates move the electron beam by electrostatic attraction and repulsion, horizontally across the CRT screen.

Aquadag coating (8). This coating performs the function of eliminating the effects of secondary emission. The aquadag is positively charged and attracts secondary emission away from the screen.

Screen (9). The screen is also called the *face*. It is coated with phosphor. When struck by electrons, the phosphor coating becomes luminous and starts glowing, thus enabling the tube to visually present electronic signals.

Figure 3.10 shows the formation of an electron beam with the controls. Astigmatism control is used to make the spot round by adjusting the potential between the deflection plates and the accelerating anode.

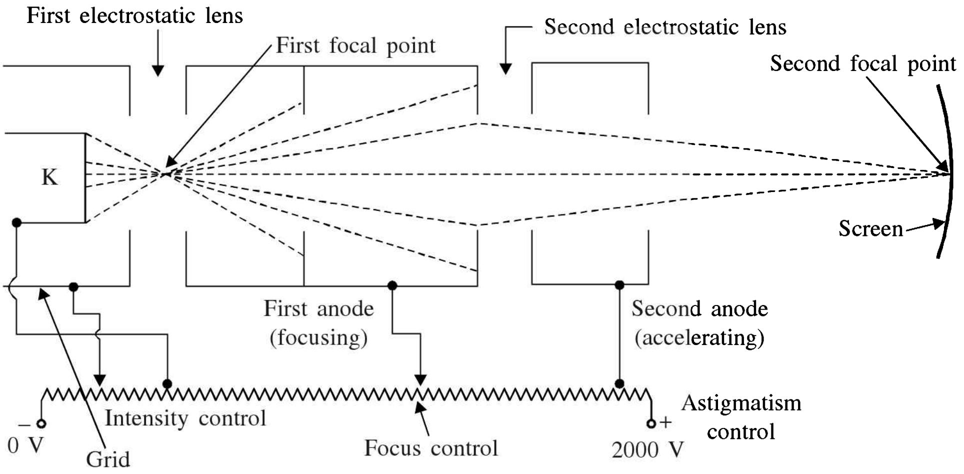


Figure 3.10 Formation of an electron beam.

The deflection on the screen (in metres) is given by

$$D = \frac{L l_d E_d}{2 d E_a}$$

where

L is the distance from the centre of the deflection plates to screen, in meters
l_d is the effective length of the deflection plates, in metres
d is the distance between the deflection plates, in metres
E_d is the deflection voltage, in volts
E_a is the accelerating voltage, in volts.

Figure 3.11 illustrates the effect of each parameter on the deflection of the beam.

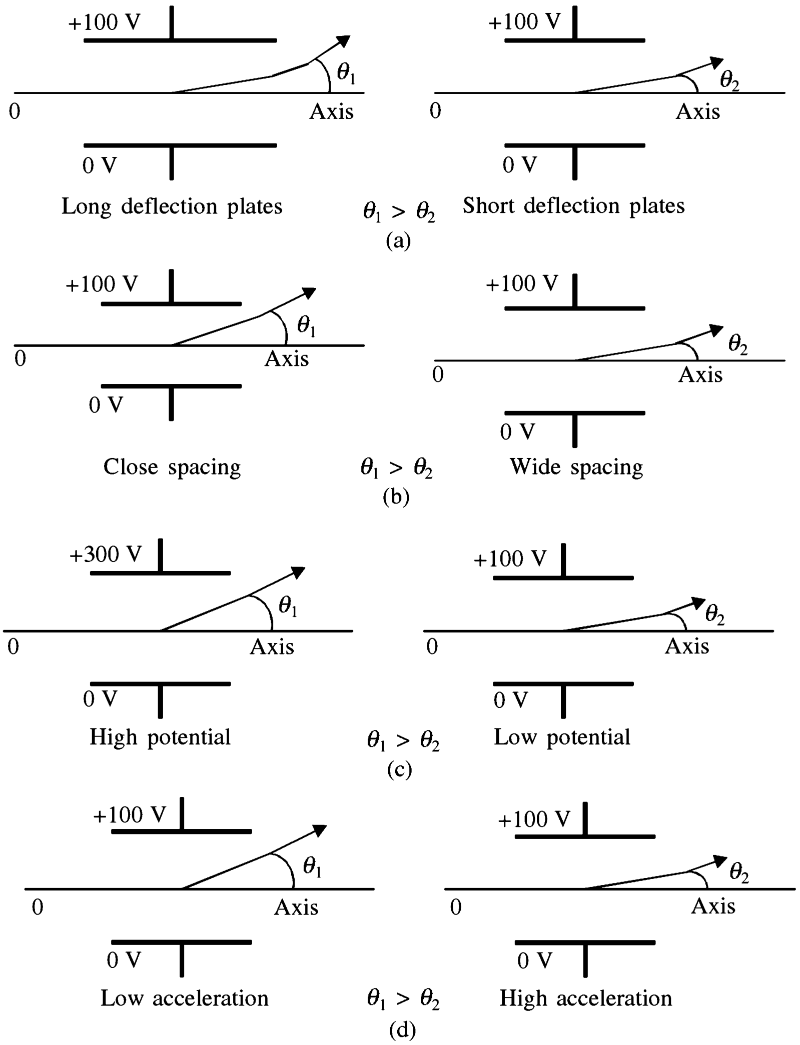


Figure 3.11 Factors influencing the deflection.

Sometimes, we define the sensitivity of CRT as follows:

$$S = \frac{D}{E_d} = \frac{Ll_d}{2dE_a}$$

Analog CRTs are vector displays that can move the beam to any point directly. A signal from the vertical amplifier moves the beam in the vertical direction. The CRT and its drivers must be able to deflect the beam vertically as fast as the signal rises. That is, the CRT bandwidth must be the same as the input bandwidth of the scope. High bandwidth CRTs pose several problems. As CRT bandwidth goes up, the following happens:

- Cost of the CRT goes up
- Accuracy of the CRT goes down
- Reliability of the CRT goes down

To keep the cost of the CRT low while still keeping the accuracy and reliability high, the scope must use as low a bandwidth CRT as possible. However, since the CRT must have the same bandwidth as the scope, high bandwidth analog scopes demand high bandwidth CRTs.

Commonly used phosphor materials

Phosphor is the material that is coated inside the CRT screen. It absorbs the kinetic energy of the electrons bombarding the material and re-emits energy in the visible spectrum. This property of the material is called *fluorescence*. Phosphor also has the property that it continues to emit light for some time even after removal of the source of excitation. This property is called *phosphorescence*. The duration of time during which phosphorescence occurs is called *persistence* of the phosphor. Persistence is measured in terms of time required for the image to decay to a certain percentage of the original light output. The intensity of the light emitted is called *luminance*. It depends upon the number of electrons bombarding the material, the energy of the electrons, the time for which electrons strike and the physical characteristics of the phosphor. Table 3.1 gives the characteristics of some of the phosphor materials used in different applications.

Table 3.1 Phosphor materials and their characteristics

Phosphor type	*Fluorescence*	*Phosphorescence*	*Relative luminance* (%)	*Decay to* 0.1% (ms)
General Purpose				
P1	Yellowish-Green	Yellowish-Green	45	95
P20	Yellowish-Green	Yellowish-Green	85	11
P31	Green	Green	100	32
Photography Recording				
P5	Blue	Blue	3	0.1
P11	Purplish-Blue	Purplish-Blue	25	17
P35	Blue-White	Blue-White	55	21
Short Decay				
P15	Bluish-Green	Bluish-Green	15	0.05
P16	Bluish-Purple	Bluish-Purple	0.1	–
P24	Greenish-Blue	Greenish-Blue	8	0.6

(*Contd.*)

Table 3.1 Phosphor material and their characteristics (*Contd.*)

Phosphor type	*Fluorescence*	*Phosphorescence*	*Relative luminance* (%)	*Decay to* 0.1% (ms)
Video Display				
P3	Green-Yellow	Green-Yellow	45	74
P4	White	White	50	20
P6	White	White	70	18
P18	White	White	18	50
P23	White	White	80	22
Long Decay—Sulfide				
P2	Blue-Green	Yellowish-Green	60	51
P7	Blue-White	Yellowish-Green	45	66
P14	Purplish-Blue	Orange	40	55
P17	Yellowish-Green	Yellowish-Green	30	82
P28	Yellowish-Green	Yellowish-Green	50	132
P32	Blue-Green	Yellowish-Green	25	235
Long Decay—Fluoride				
P12	Orange	Orange	18	700
P19	Orange	Orange	25	–
P25	Yellowish-Orange	Yellowish-Orange	12	–
P26	Orange	Orange	17	–
P33	Orange	Orange	20	–
Medium Decay—Red				
P13	Reddish-Orange	Reddish-Orange	4	375
P27	Reddish-Orange	Reddish-Orange	20	150
Variable Decay				
P34	Blue-Green	Green	17	50

Z modulation

Brilliance or intensity modulation (also called Z modulation) is obtained by the action of a potential applied to the cathode or grid that controls the intensity of the beam. Figure 3.12 shows the two techniques used for Z modulation.

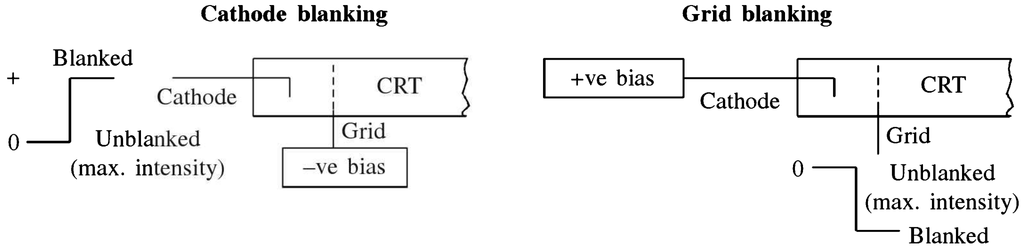


Figure 3.12 Z modulation techniques.

Generally, a change of 5 V will produce a noticeable change in brightness, while a swing of about 50 V will extinguish a maximum intensity trace. The beam is normally extinguished during *flyback* or retrace, by means of an auxiliary *blanking* electrode, which can deflect the beam so that it no longer passes through the deflection plates, and hence does not reach the screen.

Z modulation is used in applications like echo sounders, ultra sound, large radar displays, generating markers, etc. Figure 3.13 shows the output and reflected radar pulses.

Objects closer to output wave produce more intense reflected pulses. This return pulse is used for *Z* Modulation.

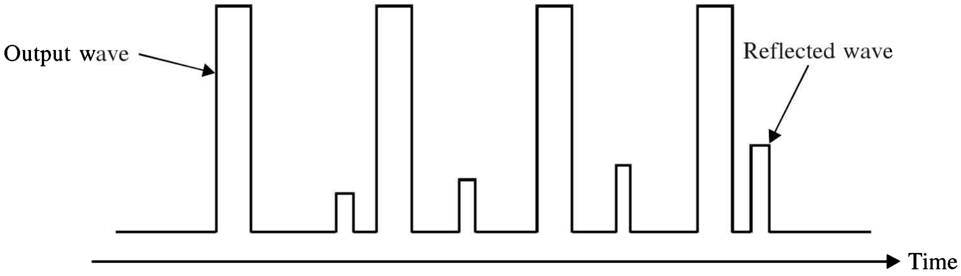


Figure 3.13 Output and reflected waves.

Graticule

There are grid markings on the screen that create the *graticule*. Each vertical and horizontal line constitutes a *major division*. The graticule is usually laid out in an 8-by-10 division pattern. Labeling on the oscilloscope controls (such as volts/div and sec/div) always refers to major divisions. The tick marks on the centre horizontal and vertical graticule lines (see Figure 3.14) are called minor divisions. Many oscilloscopes display on the screen the number of volts each vertical division represents and the number of seconds each horizontal division represents. Some scopes have provision to illuminate the graticule to distinctly see the markings.

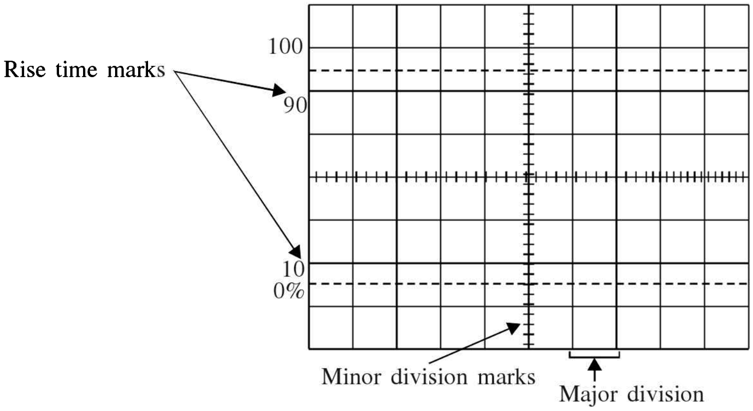


Figure 3.14 An oscilloscope graticule.

Random and raster scan

A **random scan** display device using CRT directs the electron beam only to the parts of the screen where a picture is to be drawn. This kind of device draws only a line on the screen given by its two end points. That is why these devices are referred to as vector displays (or calligraphic displays).

Consider a triangle that is to be drawn on a random scan display device as shown in Figure 3.15. Each edge of the triangle can be drawn in any order. This drawing will disappear in (for instance) 40 microseconds depending on the persistence of the phosphor coated inside the screen. So we have to refresh the screen (keep drawing the triangle before it fades) to get a picture on the screen for some time.

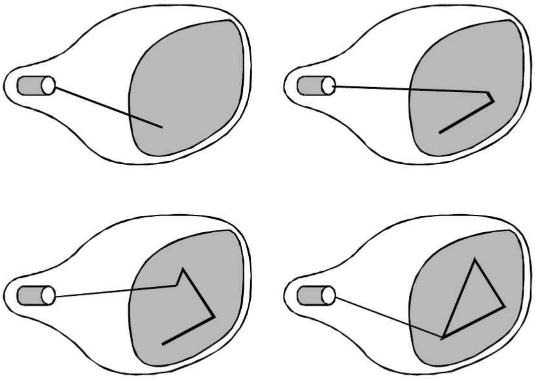


Figure 3.15 Drawing a triangle on a random scan device.

Objects are defined by a set of lines. When the object consists of too many lines, refreshing time may be longer than the persistence of the phosphor. This leads to flickering of the picture on screen. Moreover, it is almost impossible to create images with shaded objects or areas filled with a given colour.

A **raster scan** display on the other hand directs the electron beam across the screen—one row at a time from top to bottom. The screen is assumed to consist of a two dimensional matrix. Each point of the screen corresponding to a matrix element is called a *pixel* (picture element).

Consider the same triangle that is now to be drawn on a raster scan display device. This time the electron beam sweeps the whole screen, one scan-line at a time from top to bottom as shown in Figure 3.16. The pixels through which the line passes through are set to the desired colour. The lines are not straight any more as in random scan display device. They now consist of pixels. The refresh process must also be performed for raster scan monitors. Most television monitors are raster scan display devices: one scan-line at a time.

Finding the pixels through which a given line passes is called *rasterization*. Raster scan systems use a memory buffer called frame buffer (or refresh buffer) in which the intensities of the pixels are stored. The screen is refreshed using the information (intensity of each pixel) stored in the frame buffer. We may think of the frame buffer as a two-dimensional array. Each

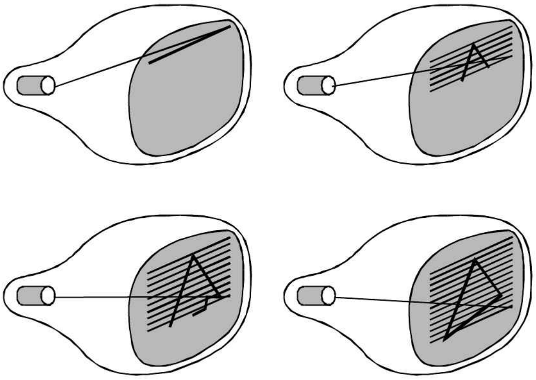


Figure 3.16 Drawing a triangle on a raster scan device.

element of the array keeps the intensity of the pixel corresponding to that element. Rasterization finds the intensity values (black or white for monochrome displays) of the pixels and sets the corresponding elements (memory locations) of the 2D array. Therefore, once the rasterization is performed, the refreshing process reads the frame buffer and turns the electron beam on or off depending on the intensity stored in the frame buffer.

The entire contents of the frame buffer should be read and the pixels activated appropriately at a sufficiently high rate to avoid flickering of the picture. Depending on the persistence of the phosphor coated inside the screen, each monitor requires a different refresh rate. In some raster display devices, a frame buffer is displayed in two passes. This kind of refreshing is called *interlaced.* In the first pass the odd rows are refreshed, and in the second pass the even rows are refreshed. Non-interlaced displays process the pixels at twice the rate of the interlaced displays. TV sets are based on the interlacing procedure.

The other property of raster display devices is the number of pixels that can be shown on the screen. This is called the *resolution.* On a display device with a resolution of 1024 × 768, there are 768 rows (scan-lines) and in each scan line there are 1024 pixels. Aspect ratio, on the other hand, is the ratio of horizontal to vertical pixels.

Raster scan displays are used in digital oscilloscopes.

3.4 DIGITAL STORAGE OSCILLOSCOPE (DSO)

Some of the systems that make up DSOs are the same as those in analog oscilloscopes; however, digitizing oscilloscopes contain additional data processing systems (see Figure 3.17.) With the added systems, the digitizing oscilloscope collects data for the entire waveform and then displays it.

The first (input) stage of a DSO is a vertical amplifier, just like the analog scopes. Vertical attenuation controls allow us to adjust the amplitude range of this stage. Next, the analog-to-digital converter (ADC) in the acquisition system samples the signal at discrete points in time and converts the signal's voltage at these points to digital values called *sample points.* The horizontal system's sample clock determines how often the ADC takes a sample. The rate at which the sample is taken is called the *sample rate* and is expressed in samples per second.

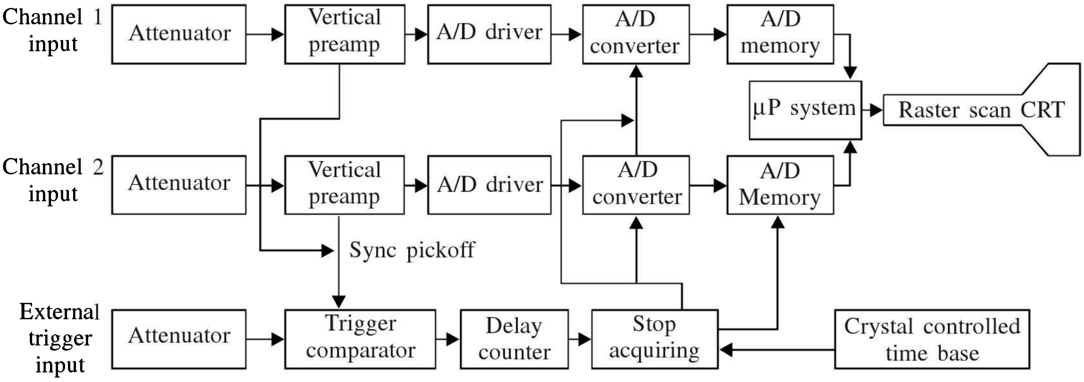


Figure 3.17 Digitizing oscilloscope—block diagram.

The sample points from the A/D converter are stored in memory as waveform points. More than one sample point may make up one waveform point. Together, the waveform points make up one waveform record. The number of waveform points used to make a waveform record is called the *record length*. The trigger system determines the start and stop points of the record. The display receives these record points after being stored in memory.

Depending on the capabilities of the oscilloscope, additional processing of the sample points may take place, enhancing the display. Pre-trigger may be available, allowing us to see events before the trigger point.

Note that the DSO's signal path includes a microprocessor (µp). The measured signal passes through this device on its way to the display. In addition to processing the signal, the microprocessor coordinates display activities, manages the front panel controls, and more. This is known as a *serial processing* architecture.

DSOs and DPOs use either an electromagnetically deflected monitor made of glass or ceramic or an LCD display. Conventional CRTs as used in analog CROs are electro-statically deflected.

Electromagnetic CRT

In modern indicator systems, electromagnetic deflection of the CRT electron beam is preferred to electrostatic deflection due to the following reasons:

(i) Increased control of the beam
(ii) Improved deflection sensitivity
(iii) Better beam position accuracy
(iv) Simpler construction of the CRT.

The primary difference between electromagnetic and electrostatic cathode ray tubes lies in the method of controlling deflection and focusing of the electron beam. Both types employ electron guns and use electrostatic fields to accelerate and control the flow of electrons. The physical construction of a CRT employing electromagnetic deflection is shown in Figure 3.18.

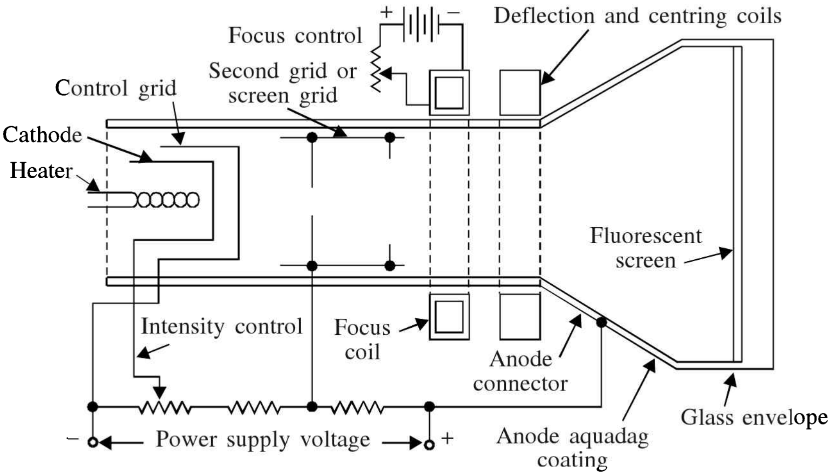


Figure 3.18 Electromagnetic CRT construction.

The electron gun in Figure 3.18 is made up of a heater, cathode, control grid, second or screen grid, focus coil, and anode (composed of a special coating). Focusing of the electron beam on the face of the screen is accomplished by the focus coil. A direct current through the windings sets up a strong magnetic field at the centre of the coil. Electrons move precisely along the axis of the tube and pass through the focusing field with no deflection. This is because they move parallel to the magnetic field at all times. Any electron, which enters the focusing field at an angle to the axis of the tube, has a force exerted on it that is perpendicular to its direction of motion. A second force on this electron is perpendicular to the magnetic lines and is, therefore, constantly changing in direction. These forces cause the electron to move in a helical or corkscrew path as shown in Figure 3.19. With the proper velocity of the electron and strength of the magnetic field, the electron will be caused to move at an angle which allows it to converge with other electrons at some point on the CRT screen. Focusing is accomplished by adjusting the current flow through the focusing coils.

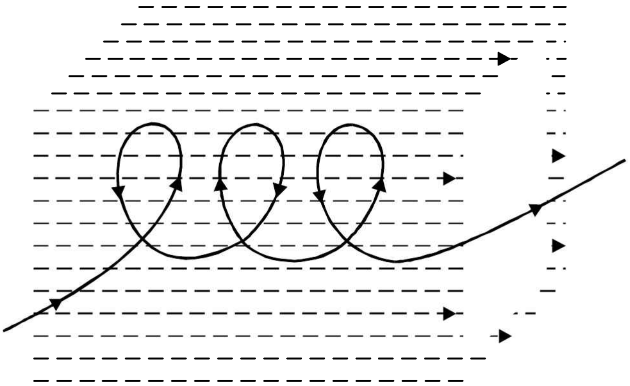
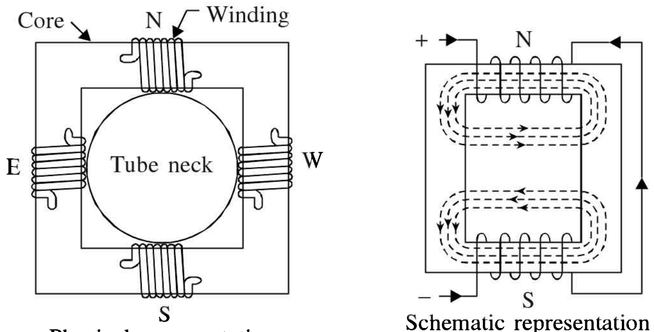


Figure 3.19 Helical motion of an electron through a uniform magnetic field.

The focused electron beam is deflected by a magnetic field that is generated by current flow through a set of deflection coils, as shown in Figure 3.18. These coils are mounted around the outside surface of the neck of the CRT. Normally, four deflection coils (N, S, E, and W) are used, as shown in Figure 3.20. Two coils in series are positioned in a manner that causes the magnetic field produced to be in a vertical plane. The other two coils, also connected in series, are positioned so that their magnetic field is in a horizontal plane. The coils (N-S) which produce a horizontal field are called the *vertical deflection* coils and the coils (E-W) which produce a vertical field are called the *horizontal deflection* coils. This may be more clearly understood if we recall that an electron beam will be deflected at right angles to a deflecting field. The deflection coils are illustrated in Figure 3.20(a). Figure 3.20(b) shows the N-S windings in schematic form.



Physical representation

Schematic representation

Figure 3.20 Deflection yoke.

Electron deflection in the electromagnetic CRT is proportional to the strength of the magnetic fields. Magnetic field strength depends on the current in the coils. The sweep circuits associated with electromagnetically deflected CRTs must provide currents rather than voltage to produce the desired beam deflection.

Sampling methods

Digitizing oscilloscopes, DSO or DPO, can use either real-time, interpolated real-time, or equivalent-time sampling to collect sample points. Real-time sampling is ideal for signals whose frequency is less than half the scope's maximum sample rate. Here, the oscilloscope can acquire more than enough points in one "sweep" of the waveform to construct an accurate image (see Figure 3.21). Note that real-time sampling is the only way to capture single-shot transient signals with a digitizing scope.

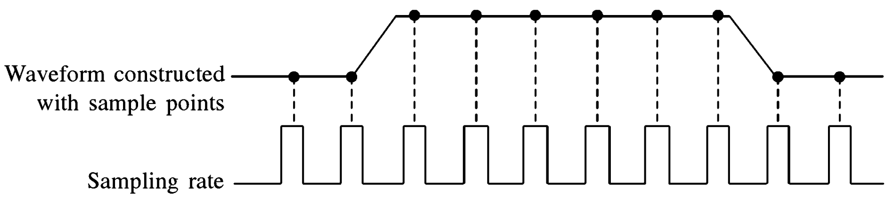


Figure 3.21 Real-time sampling.

When measuring high-frequency signals, the oscilloscope may not be able to collect enough samples in one sweep. There are two solutions for accurately acquiring signals whose frequency exceeds half the scope's sample rate:

(a) Collect a few sample points of the signal in a single pass (in real-time mode) and use interpolation to fill in the gaps. Interpolation is a processing technique to estimate what the waveform looks like based on a few points.

(b) Build a picture of the waveform by acquiring samples from successive cycles of the waveform, assuming the signal repeats itself (equivalent-time sampling mode).

Real-time sampling with interpolation

Digitizing oscilloscopes take discrete samples of the signal which may be displayed. However, it may be difficult to visualize the signal represented as dots, especially because there may be only a few dots representing high frequency portions of the signal. To aid in the visualization of signals, digitizing oscilloscopes typically have interpolation display modes.

In simple terms, interpolation *connects the dots*. Using this process, a signal that is sampled only a few times in each cycle can be accurately displayed. However, for accurate representation of the signal, the sample rate should be at least four times the bandwidth of the signal.

Linear interpolation connects sample points with straight lines. This approach is limited to reconstructing straight-edged signals like square waves. The more versatile sin *x*/*x* interpolation connects sample points with curves (see Figure 3.22). The interpolation sin *x*/*x* is a mathematical process in which points are calculated to fill in the time between the real samples.

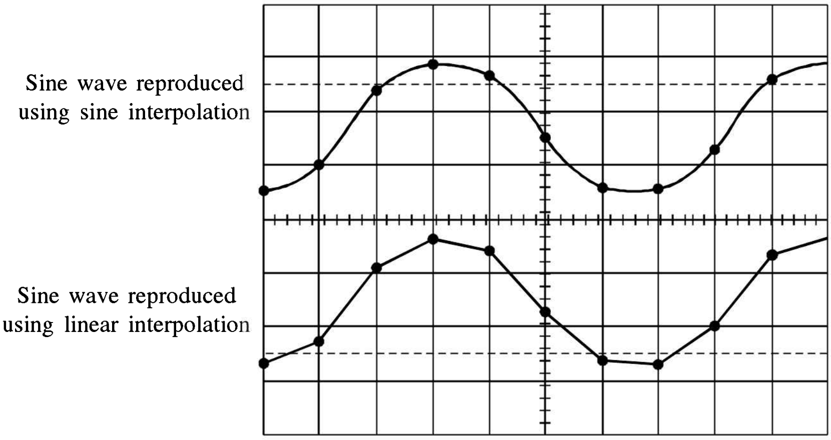

Sine wave reproduced using sine interpolation
Sine wave reproduced using linear interpolation

Figure 3.22 Linear and sine interpolation.

This form of interpolation lends itself to curved and irregular signal shapes, which are far more common in the real world than pure square waves and pulses. Consequently, sin *x*/*x* interpolation is the most preferred method for most applications.

Equivalent-time sampling

Some digitizing oscilloscopes can use equivalent-time sampling to capture very fast repeating signals. Equivalent-time sampling constructs a picture of a repetitive signal by capturing a little bit of information from each repetition (see Figure 3.23). The waveform slowly builds up like a string of lights going on one-by-one. With sequential sampling the points appear from left to right in sequence; with random sampling the points appear randomly along the waveform.

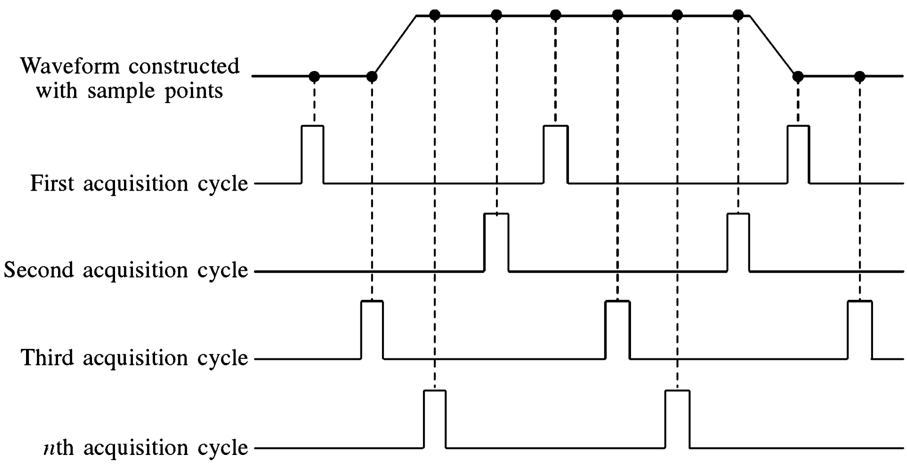


Figure 3.23 Equivalent-time sampling.

3.5 DIGITAL PHOSPHOR OSCILLOSCOPE (DPO)

The DPO architecture is similar to that of DSO but has special features designed to re-create the intensity grading of an analog CRT. Rather than relying on a chemical phosphor as an analog scope does, the DPO has a purely electronic digital phosphor which is actually a continuously updated database. This database has a separate *cell* of information for every single pixel in the scope's display. Each time a waveform is captured (in other words, every time the scope triggers) it is mapped into cells of the digital phosphor database. Each cell representing a screen location that is touched by the waveform, gets reinforced with intensity information; others do not. Thus intensity information builds up in cells where the waveform passes most often. When the digital phosphor database is fed to the oscilloscope's display, the display reveals intensified waveform areas, in proportion to the signal's frequency of occurrence at each point, much like the intensity grading characteristics of an analog oscilloscope. Further, unlike an analog scope, the DPO allows the varying levels to be expressed in contrasting colours if desired. With a DPO, it is easy to see the difference between a waveform that occurs on almost every trigger and one that occurs, say, every 100th trigger.

Figure 3.24 shows the changes in the vertical input architecture of DSO to achieve all this manipulation without slowing down the whole acquisition process. Display gets data directly from the memory.

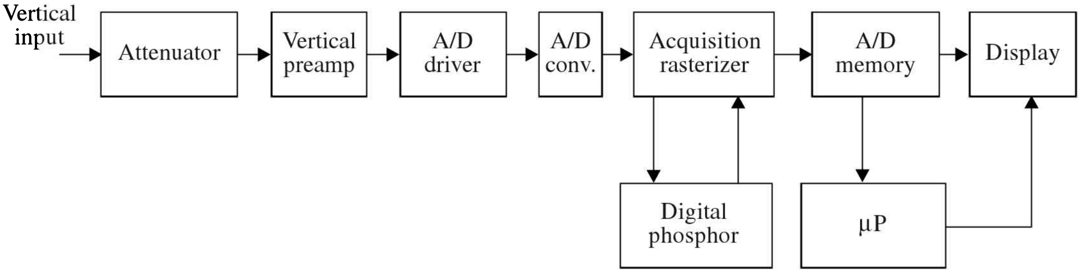


Figure 3.24 Digital phosphor oscilloscope—block diagram.

3.6 THE CONTROLS

The basic controls found on most of the analog and digitizing oscilloscopes are analysed in the following sections. Note that some controls differ between analog and digitizing oscilloscopes.

3.6.1 Display Controls

Display systems vary between analog and digitizing scopes, including both DSOs and DPOs. Common controls include:

(a) An intensity control to adjust the brightness of the waveform. As we increase the sweep speed of an analog oscilloscope, we need to increase the intensity level.

(b) A focus control to adjust the sharpness of the waveform. Digitizing oscilloscopes may not have a focus control.

(c) A trace rotation control to align the waveform trace with the screen’s horizontal axis. The position of the oscilloscope in the earth's magnetic field affects waveform alignment. Digitizing oscilloscopes may not have a trace rotation control.

(d) On DPOs, a contrast control.

(e) On many DSOs and on DPOs, a colour palette control to select trace colours and intensity grading colour levels.

(f) Other display controls may adjust the intensity of the graticule lights and turn on or off any on-screen information (such as menus).

3.6.2 Vertical Controls

Vertical controls are used to position and scale the waveform vertically. The oscilloscope also has controls for setting the input coupling and other signal conditioning, described later in this section.

Position and volts per division

The vertical position control moves the waveform up or down to exactly where we want it on the screen. The volts per division (usually written volts/div) setting varies the size of the waveform on the screen. A good general purpose oscilloscope can accurately display signal levels from about 4 mV to 40 V.

The volts/div setting is a scale factor. For example, if the volts/div setting is 5 V, then each of the eight vertical divisions represents 5 V and the entire screen can show 40 V from bottom to top (assuming a graticule with eight major divisions). If the setting is 0.5 volts/div, the screen can display 4 V from bottom to top, and so on. The maximum voltage that we can display on the screen is obtained by multiplying the volts/div setting by the number of vertical divisions. If an external probe is used for attenuation of the signal, we must divide the volts/div scale by the attenuation factor of the probe. Often the volts/div scale has either a variable gain or a fine gain control for scaling a displayed signal to a certain number of divisions. This control can be used to take rise time measurements.

Input coupling

Coupling means the method used to connect an electrical signal from one circuit to another. In this case, the input coupling is the connection from the test circuit to the oscilloscope. The coupling can be set to dc, ac, or ground. The dc coupling shows all of an input signal. The internal arrangement is shown in Figure 3.25. Because of the capacitor present, ac coupling blocks the dc component of a signal so that we see the waveform centred at 0 V. Figure 3.26 illustrates this difference. The ac coupling setting is handy when the entire signal (alternating plus constant components) is too large for the volts/div setting.

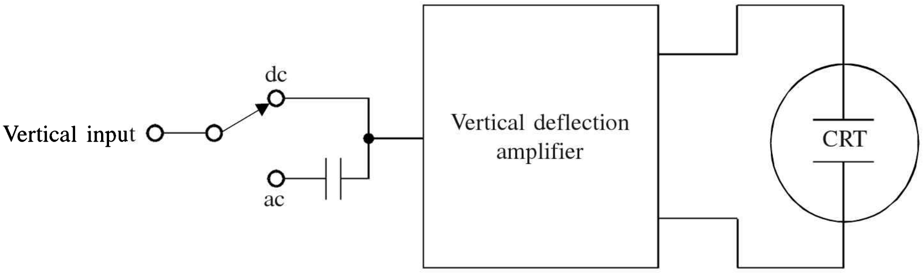
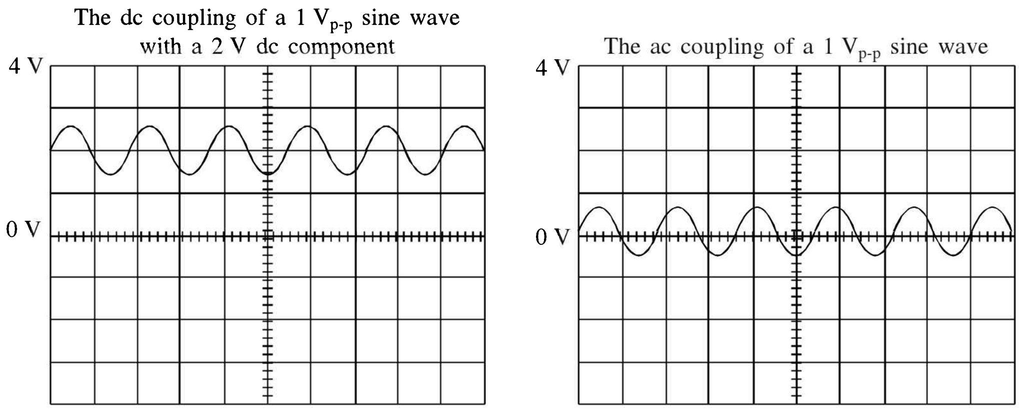


Figure 3.25 Vertical input coupling arrangement.


The dc coupling of a 1 $V_{p\text{-}p}$ sine wave with a 2 V dc component

The ac coupling of a 1 $V_{p\text{-}p}$ sine wave

Figure 3.26 The ac and dc input coupling.

The ground setting disconnects the input signal from the vertical system, which lets us see where 0 V is on the screen. With grounded input coupling and auto trigger mode, we see a horizontal line on the screen that represents 0 V. Switching from dc to ground and back again is a handy way of measuring signal voltage levels with respect to ground.

AC coupling may cause voltage droop due to the loss of low frequencies. Figure 3.27 illustrates this.

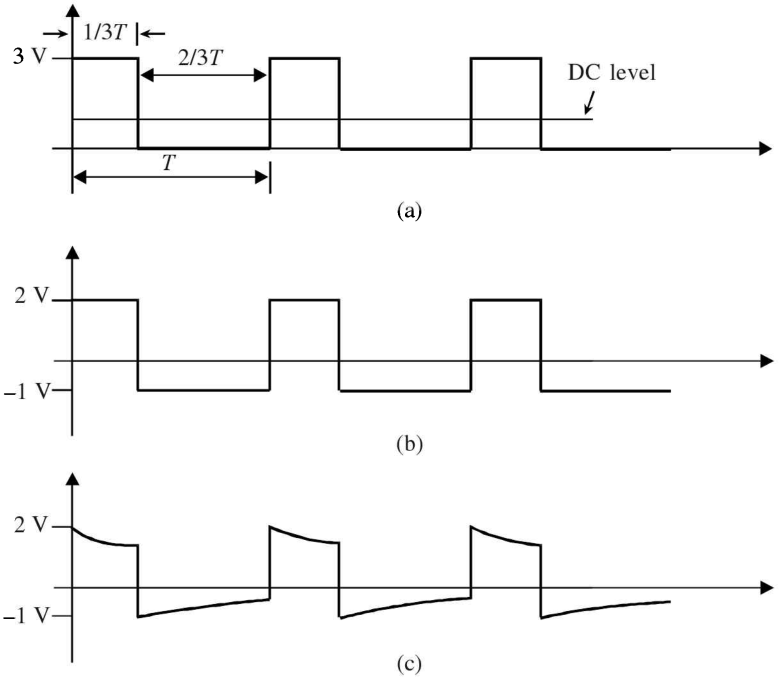


Figure 3.27 Effect of ac coupling.

Bandwidth limit

Most oscilloscopes have a circuit that limits the bandwidth of the oscilloscope. By limiting the bandwidth, we reduce the noise that sometimes appears on the displayed waveform, providing us with a more refined signal display. Figure 3.28 illustrates this. The control is called *Bandwidth Limit Control.*

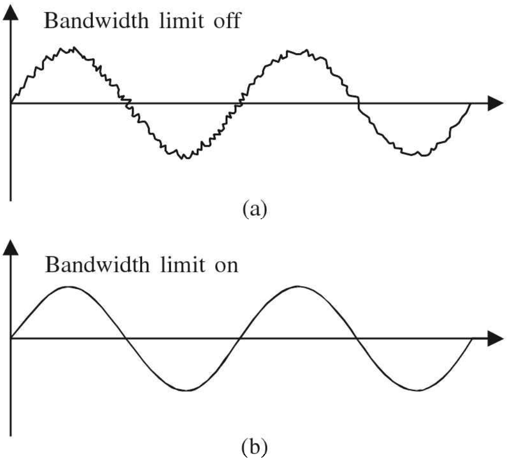
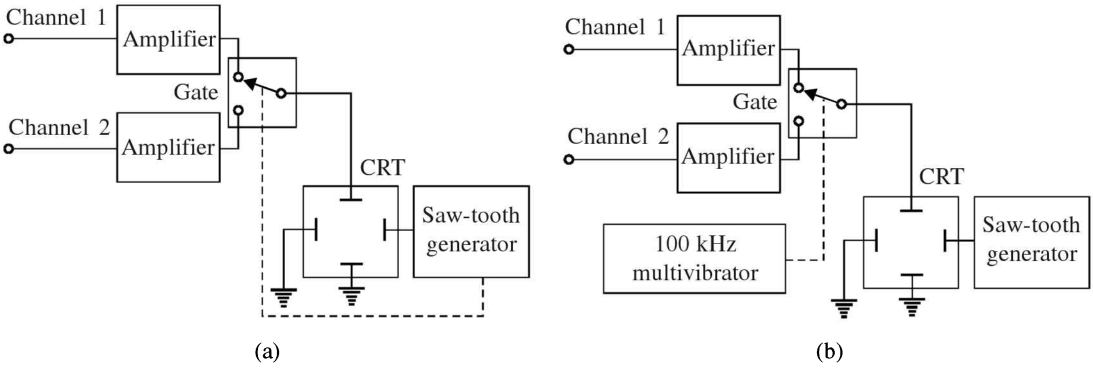


Figure 3.28 Effect of Bandwidth Limit Control.

Alternate and chop display

On analog scopes, multiple channels are displayed using either the ALT or the CHOP mode. Digitizing oscilloscopes can present multiple channels simultaneously without the need for chop or alternate modes. Figure 3.29 shows the internal arrangement.



(a) (b)

Figure 3.29 (a) ALT and (b) CHOP—internal arrangement.

The ALT mode draws each channel alternately—the oscilloscope completes one sweep on channel 1, then one sweep on channel 2, a second sweep on channel 1, and so on. This mode is used with medium- to high-speed signals, when the sec/div scale is set to 0.5 ms or faster.

The CHOP mode causes the oscilloscope to draw small parts of each signal by switching back and forth between them. The switching rate is too fast for us to notice, so the waveform looks a whole. We typically use this mode with slow signals requiring sweep speeds of 1 ms per division or less. Figure 3.30 shows the difference between the two modes. It is often useful to view the signal both ways, to make sure that we have the best view.

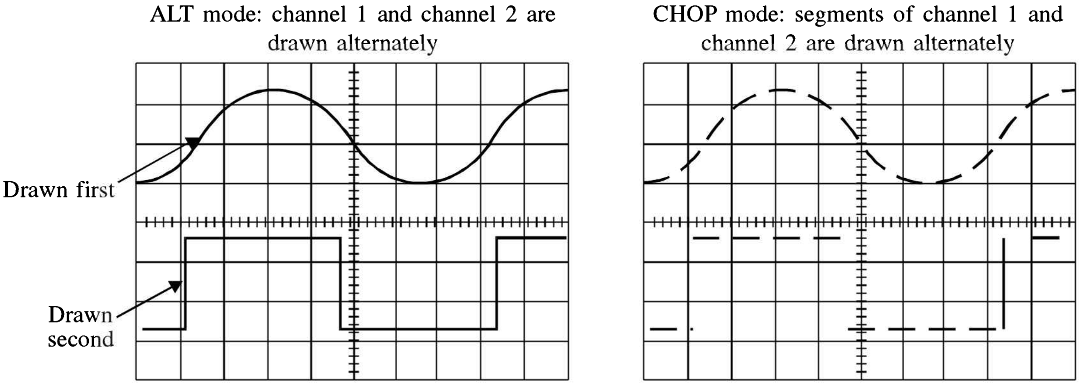

ALT mode: channel 1 and channel 2 are drawn alternately
CHOP mode: segments of channel 1 and channel 2 are drawn alternately
Drawn first
Drawn second

Figure 3.30 Multi-channel display modes.

Math operations

The oscilloscope may also have operations to add waveforms together, creating a new waveform display. The analog oscilloscopes combine the signals while the digitizing oscilloscopes create new waveforms mathematically. Subtracting waveforms is another math operation. Subtraction with analog oscilloscopes is possible by using the channel invert function on one signal and then using the add operation. Digitizing oscilloscopes typically have a subtraction operation available. Figure 3.31 illustrates a third waveform created by adding two different signals together.

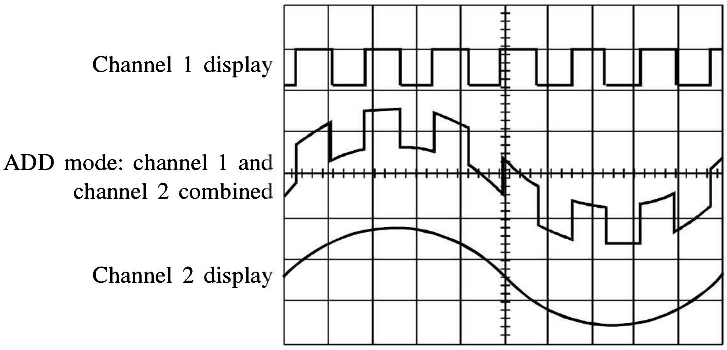


Figure 3.31 Adding channels.

Using the power of their internal processors, digitizing oscilloscopes offer many advanced math operations: multiplication, division, integration, Fast Fourier Transform, and more.

3.6.3 Horizontal Controls

Horizontal controls are used to position and scale the waveform horizontally.

Position and seconds per division

The horizontal position control moves the waveform from left and right to exactly where we want it on the screen. The seconds per division (usually written as sec/div) setting lets us select the rate at which the waveform is drawn across the screen (also known as the time base setting or sweep speed). This setting is a scale factor. For example, if the setting is 1 ms, each horizontal division represents 1 ms and the total screen width represents 10 ms (ten divisions). Changing the sec/div setting lets us look at longer or shorter time intervals of the input signal.

As with the vertical volts/div scale, the horizontal sec/div scale may have variable timing, allowing us to set the horizontal time scale in between the discrete settings.

Time base selections

The oscilloscope has a time base usually referred to as the main time base and it is probably the most useful. Many oscilloscopes have what is called a delayed time base—a time base sweep that starts after a pre-determined time from the start of the main time base sweep. Using a delayed time base sweep allows us to see events more clearly or even see events not visible with the main time base sweep alone. The delayed time base requires the setting of a delay time and possibly the use of delayed trigger modes and other settings.

Trigger position

Horizontal trigger position control is only available on digitizing oscilloscopes. The trigger position control may be located in the horizontal control section of the oscilloscope. It actually represents the *horizontal* position of the trigger in the waveform record.

Varying the horizontal trigger position allows us to capture what a signal did *before* a trigger event (called *pretrigger viewing*). Thus it determines the length of the viewable signal both preceding and following a trigger point.

Digitizing oscilloscopes can provide pretrigger viewing because they constantly process the input signal, i.e. whether a trigger has been received or not. A steady stream of data flows through the oscilloscope; the trigger merely tells the oscilloscope to save the present data in memory. In contrast, analog oscilloscopes only display the signal (that is, write it on the CRT) after receiving the trigger.

Pretrigger viewing is a valuable troubleshooting aid. For example, if a problem occurs intermittently, we can trigger on the problem, record the events that led to it and, possibly, find the cause.

Zoom

The oscilloscope may have special horizontal magnification settings that lets us display a magnified section of the waveform on the screen. On a DSO, the operation is performed on stored digitized data.

XY mode

Most analog oscilloscopes have the capability of displaying a second channel signal along the X-axis (instead of time). This is known as XY mode. To use the oscilloscope in the XY mode, one input is fed to the horizontal input and the other to the vertical input.

The Z-axis

The Z-axis brings a third dimension (intensity) to the traditional waveform display. One application of the Z-axis is to feed special timed signals into the separate Z-input to create highlighted *marker* dots at known intervals in the waveform.

XYZ mode

Analog oscilloscopes and DPOs can use the Z-input to create an XY display with intensity grading. In this case, the oscilloscope samples the instantaneous data value at the Z-input and uses that value to intensify a specific part of the waveform. One application of this mode is for displaying the polar patterns commonly used in testing wireless communication devices.

3.6.4 Trigger Controls

The trigger controls stabilize repeating waveforms and capture single-shot waveforms. The trigger makes repeating waveforms appear static on the oscilloscope display by repeatedly displaying the same portion of the input signal. Without this control, we will see a jumble on the screen that as a result of each sweep starting at a different place on the signal (see Figure 3.32).

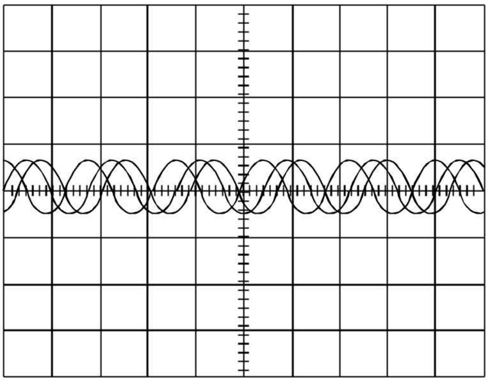


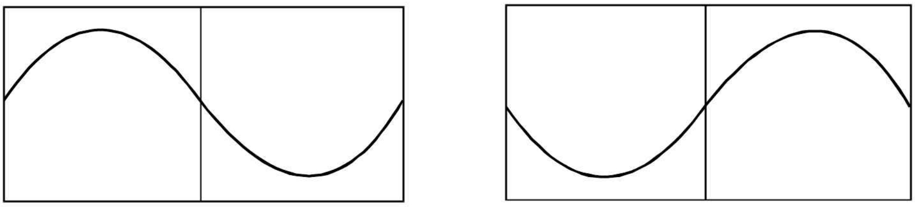
Figure 3.32 Untriggered display.

Trigger level and slope

The oscilloscope may have several different types of triggers, such as edge, video, pulse, or logic. Edge triggering is the basic and most common type.

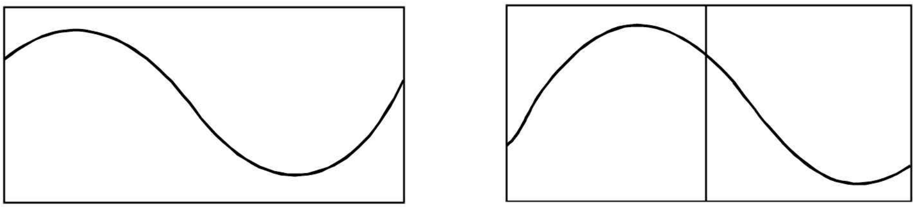
For edge triggering, the trigger level and slope controls provide the basic trigger point definition. The trigger circuit acts as a comparator. We select the slope and voltage level of one side of the comparator. When the trigger signal matches our settings, the oscilloscope generates a trigger. The slope control determines whether the trigger point is on the rising or the falling edge of a signal. A rising edge is a positive slope and a falling edge is a negative slope. The level control determines where on the edge the trigger point occurs.

Figure 3.33 defines the slope and level used for trigger control. Figure 3.34 illustrates the effect of positive and negative slope triggering. Figure 3.35 shows how the trigger is generated when a sine wave is applied to the vertical input. This sine wave is used to generate pulses as per the settings of Trigger Level and Trigger Slope. The pulse generator is set such that every time the triggering signal (here sine wave) reaches a specific level, either with positive or with negative slope (one or the other), the pulse generator outputs a pulse.



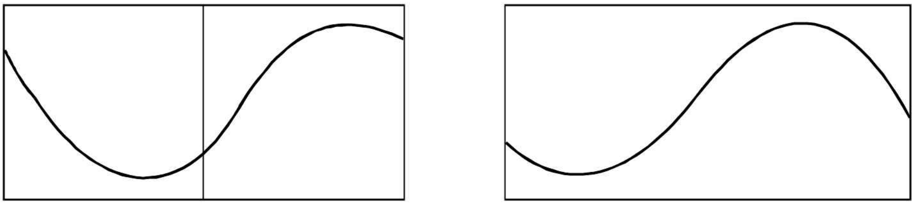
Slope = +, level = 0
(a)

Slope = –, level = 0
(b)



Slope = +, level = +
(c)

Slope = +, level = –
(d)



Slope = –, level = +
(e)

Slope = –, level = –
(f)

Figure 3.33 Defining slope and level.

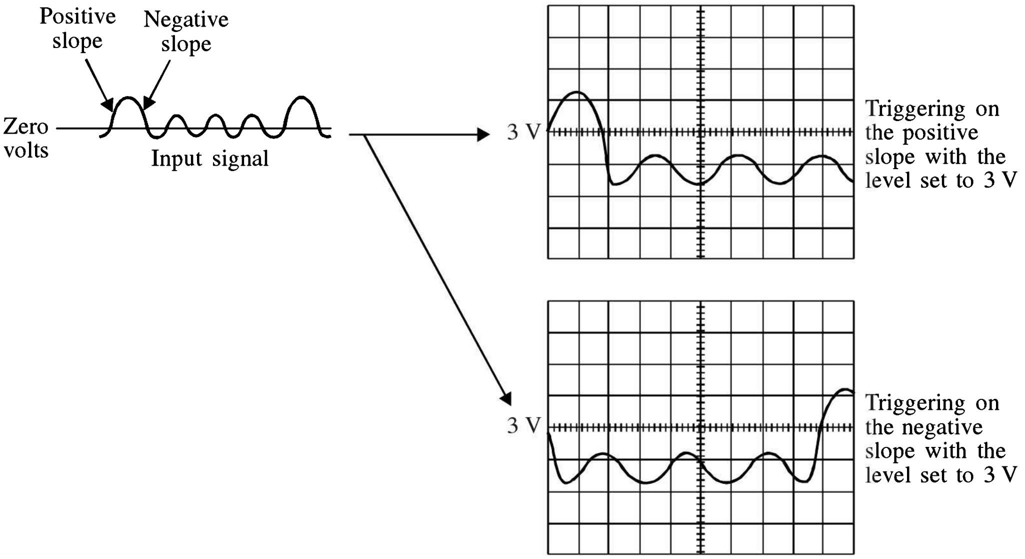

Positive slope
Negative slope
Zero volts
Input signal
3 V
Triggering on the positive slope with the level set to 3 V
3 V
Triggering on the negative slope with the level set to 3 V

Figure 3.34 Positive and negative slope triggering.

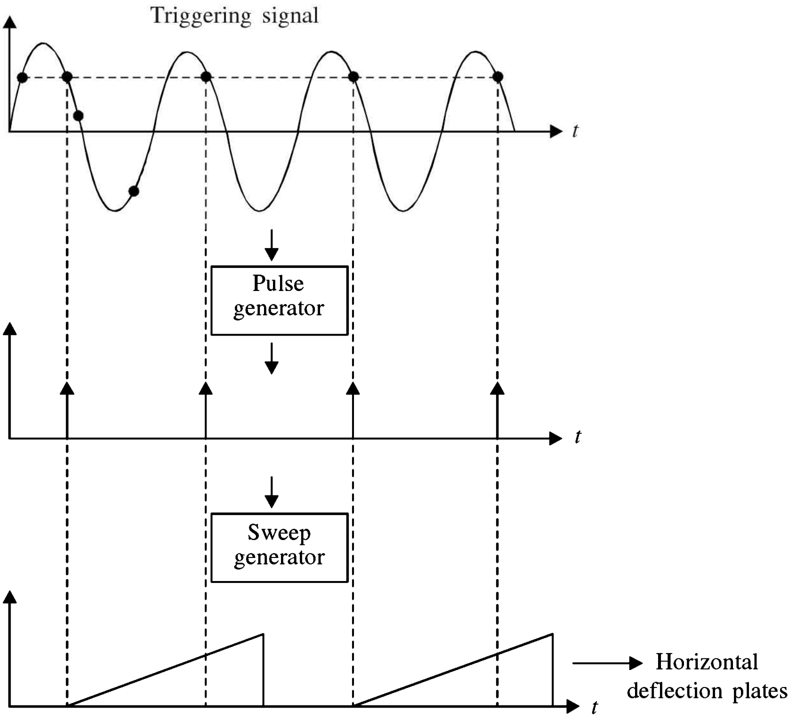
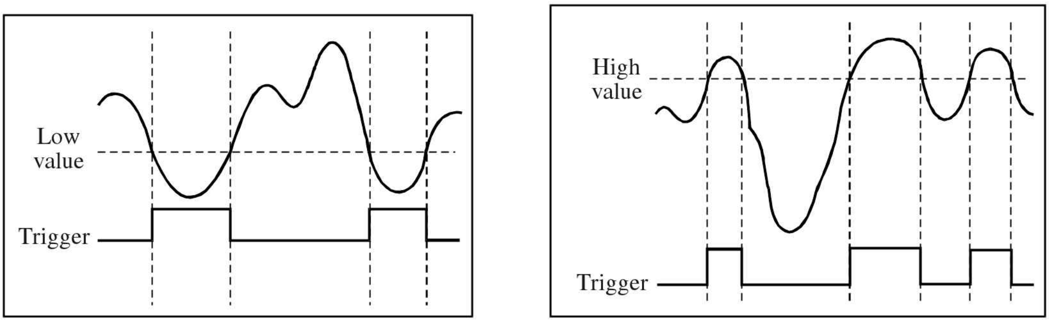
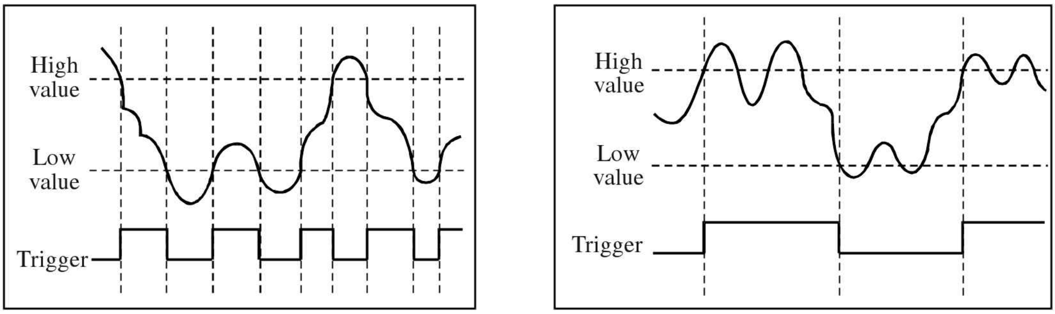


Figure 3.35 Triggering process, with the scope set to trigger on the negative slope.

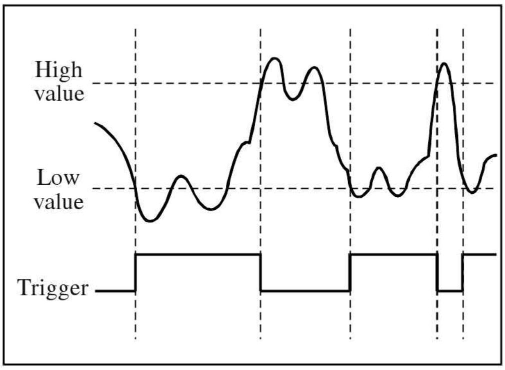
Modern oscilloscopes have a number of options available for generating triggering pulses. Figure 3.36 gives an example of these options.



Below low-level triggering Above high-level triggering



Inside region triggering High value with low value hysteresis triggering mode



Low value with high value hysteresis triggering

Figure 3.36 Various triggering methods.

Delayed triggering

The delayed trigger or delayed sweep feature creates a small time delay between the beginning of a waveform and initiation of a sweep cycle. This feature allows us to view a small section of a larger waveform, for example, some oscillations or some ringing or some noise riding on a larger waveform. Figure 3.37 illustrates this with an example where a 5 MHz signal is riding on a signal of 100 kHz frequency. The period of 5 MHz signal is 0.2 μs whereas for 100 kHz it is 10 μs. Further, 5 MHz signal starts after 5.5 μs from the triggering point. As such if we adjust the time base to see the waveform of 100 kHz, the details of 5 MHz signal will not be visible. In order to view this signal, trigger the sweep on the 100 kHz signal, adjust the trigger delay control to 5.5 μs and set the time base to 0.2 μs. Now we will see only the 5 MHz signal on the screen. The 100 kHz signal will be suppressed.

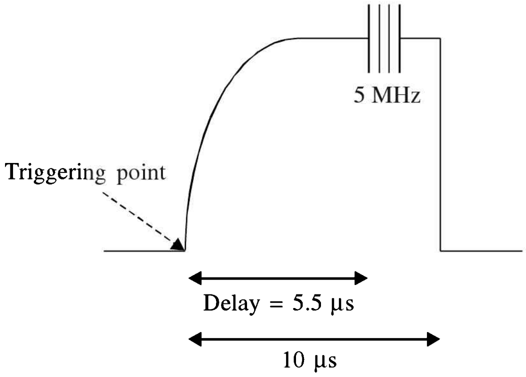


Figure 3.37 Delayed sweep or delayed triggering.

Trigger sources

The oscilloscope does not necessarily have to trigger on the signal being measured. Following sources can trigger the sweep:

- Any input channel
- An external source other than the signal applied to an input channel
- The power source signal
- A signal internally generated by the oscilloscope

Most of the time we can leave the oscilloscope set to trigger on the channel displayed. Many oscilloscopes provide a trigger *output* that delivers the trigger signal to another instrument.

Note that the oscilloscope can use an alternate trigger source whether displayed or not. So we have to be careful not to trigger on, for example, channel 1 while displaying channel 2.

Trigger modes

The trigger mode determines whether or not the oscilloscope draws a waveform if it does not detect a trigger. Common trigger modes include *normal* and *auto*.

In *normal* mode the oscilloscope only sweeps if the input signal reaches the set trigger point; otherwise (on an analog oscilloscope) the screen is blank or (on a digitizing oscilloscope) frozen on the last acquired waveform. Normal mode can be disorienting since we may not see the signal at first if the level control is not adjusted correctly.

Auto mode causes the oscilloscope to sweep, even without a trigger. If no signal is present, a timer in the oscilloscope triggers the sweep. This ensures that the display will not disappear if the signal drops to small voltages. It is also the best mode to use if we are looking at many signals and do not want to set the trigger each time.

In practice, we will probably use both modes—normal mode because it lets us select just the signal area we need to see, and auto mode because it requires less adjustment. Some oscilloscopes also include special modes for single sweeps, triggering on video signals, or automatically setting the trigger level.

Trigger coupling

Just as we can select either ac or dc coupling for the vertical system, we can choose the kind of coupling for the trigger signal. Besides ac and dc coupling, the oscilloscope may also have high frequency rejection, low frequency rejection, and noise rejection trigger coupling. These special settings are useful for eliminating noise from the trigger signal to prevent false triggering.

Trigger holdoff

Sometimes getting an oscilloscope to trigger on the correct part of a signal requires great skill. Many oscilloscopes have special features to make this task easier.

Trigger *holdoff* is an adjustable period of time during which the oscilloscope cannot trigger. This feature is useful when we are triggering on complex waveform shapes, so that the oscilloscope only triggers on the first eligible trigger point. Figure 3.38 shows how using trigger holdoff helps create a usable display.

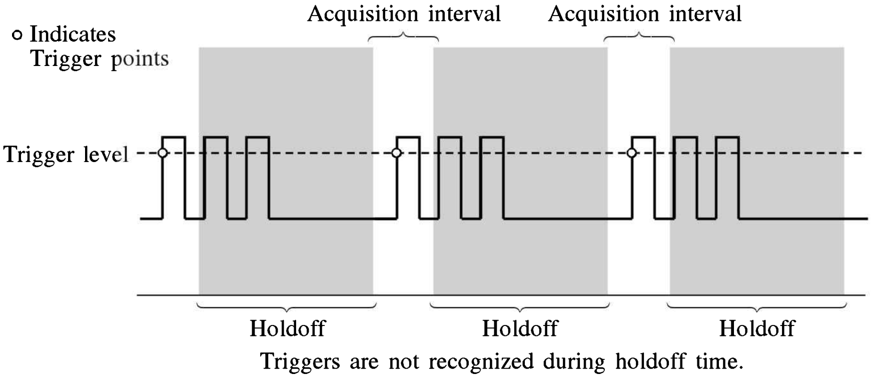

Acquisition interval
Acquisition interval
o Indicates Trigger points
Trigger level
Holdoff
Holdoff
Holdoff
Triggers are not recognized during holdoff time.

Figure 3.38 Trigger holdoff.

Special trigger settings

In addition to the usual threshold triggering, many digitizing oscilloscopes offer a host of specialized trigger settings which have no equivalents on analog instruments. These triggers respond to specific conditions in the incoming signal, making it easy to detect, for example, a pulse that is narrower than it should be. Such a condition would be impossible to detect with a voltage threshold trigger alone. Following is a partial list of the digital triggers found on advanced DSOs and DPOs, along with a brief definition of each:

Pulse width and glitch trigger. Detects pulses either within or exceeding specified widths.

Runt pulse trigger. Detects a pulse that crosses the smaller but not the greater of the two threshold levels.

Logic (Boolean) trigger. Uses multiple oscilloscope inputs as binary inputs that must meet logical conditions such as NAND or NOR to produce a trigger.

Serial data trigger. Detects specific data combinations in digital telecom signals.

Setup and hold violation trigger. Detects violations of digital setup and hold time when clock and data signals are acquired on two different inputs.

3.6.5 Acquisition Controls for Digitizing Oscilloscopes

Digitizing oscilloscopes have settings that let us control how the acquisition system processes a signal.

Acquisition modes

Acquisition modes control how waveform points are produced from sample points. Recall that sample points are the digital values that come directly out of the analog-to-digital-converter (ADC). The time between sample points is called the *sample interval*. Waveform points are the digital values that are stored in memory and displayed to form the waveform. The time difference between the waveform points is called the *waveform interval*. The sample interval and the waveform interval may or may not be the same. This leads to the existence of several different acquisition modes in which one waveform point is made up from several sequentially acquired sample points. Additionally, waveform points can be created from a composite of samples taken from multiple acquisitions, which leads to another set of acquisition modes. A description of the most commonly used acquisition modes follows:

Sample mode. This is the simplest acquisition mode. The oscilloscope creates a waveform point by saving one sample point during each waveform interval.

Peak detect mode. The oscilloscope saves the minimum and maximum value sample points taken during the two waveform intervals and uses these samples as the two corresponding waveform points. Digitizing oscilloscopes with the peak detect mode run the ADC at a fast sample rate, even at very slow time-base settings (long waveform interval), and are able to capture fast signal changes that would occur between the waveform points if in sample mode. Peak detect mode is particularly useful for seeing narrow pulses spaced far apart in time.

Hi Res mode. Like peak detect, the Hi Res mode is a way of getting more information in cases when the ADC can sample faster than the time-base setting requires. In this case, multiple samples taken within one waveform interval are averaged together to produce one waveform point. The result is a decrease in noise and an improvement in resolution for low-speed signals.

Envelope mode. Envelope mode is similar to peak detect mode. However, in envelope mode, the minimum and maximum waveform points from multiple acquisitions are combined to form a waveform that shows min/max changes over time. Peak detect mode is usually used to acquire the records that are combined to form the envelope waveform.

Average mode. In average mode, the oscilloscope saves one sample point during each waveform interval as in sample mode. However, waveform points from consecutive acquisitions are then averaged together to produce the final displayed waveform. Average mode reduces noise without loss of bandwidth but requires a repeating signal.

Stopping and starting the acquisition system

One of the greatest advantages of digitizing oscilloscopes is their ability to store waveforms for later viewing. To this end, there are usually one or more controls on the front panel that allow us to stop and start the acquisition system so that we can analyze waveforms at leisure. Additionally, we may want the oscilloscope to automatically stop acquiring after one acquisition is complete or after one set of records has been turned into an envelope or average waveform. This feature is commonly called *single sweep* or *single sequence* and its controls are usually found either with the other acquisition controls or with the trigger controls.

3.6.6 Sampling Methods

In digitizing oscilloscopes that use either real-time sampling or equivalent-time sampling as described earlier, the acquisition controls will allow us to choose which mode to use for acquiring signals. Note that this choice makes no difference for slow time-base settings, and only has an effect when the ADC cannot sample fast enough to fill the record with waveform points in one pass.

3.6.7 Other Controls

The oscilloscope may have other controls for various functions. Some of these may include:

- Measurement cursors
- Keypads for mathematical operations or data entry
- Printing capabilities
- Interfaces for connecting oscilloscope to a computer

The oscilloscope's manual normally gives details of these controls.

3.7 MEASUREMENTS

The two most basic measurements that we can make are voltage and time measurements. Every other measurement is based on one of these two fundamental techniques. We will discuss methods of taking measurements visually with the oscilloscope screen. This is a common technique with analog instruments, and also may be useful for interpretation of DSO or DPO displays. Most digitizing oscilloscopes include automated measurement tools. Knowing how to make measurements manually as described here will help us understand and check the automatic measurements of DSOs and DPOs. Automated measurements are explained in Section 3.7.4.

3.7.1 Voltage Measurements

The oscilloscope is primarily a voltage-measuring instrument. Once we have measured the voltage, other quantities can be calculated. Figure 3.39 shows the peak voltage, V_p, the peak-to-peak voltage, V_{p-p}, and the rms voltage, V_{rms}.

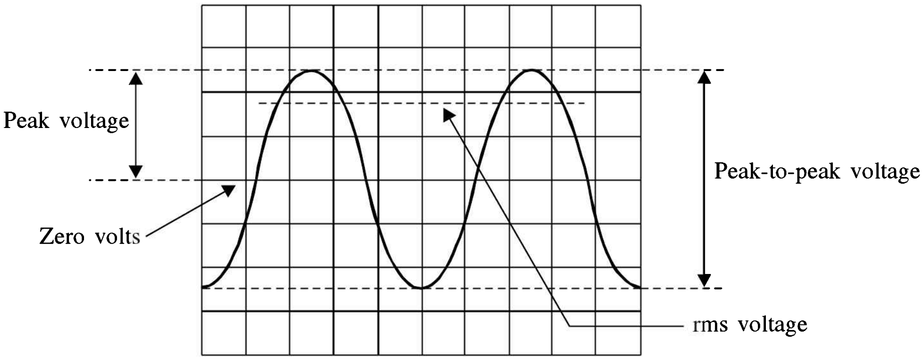


Figure 3.39 Peak voltage and peak-to-peak voltage.

The most basic method of taking voltage measurements is to count the number of divisions a waveform spans on the oscilloscope's vertical scale. Adjusting the signal to cover most of the screen vertically, then taking the measurement along the centre vertical graticule line having the smaller divisions, makes for the best voltage measurements (see Figure 3.40). The more screen area we use, the more accurately we can read from the screen.

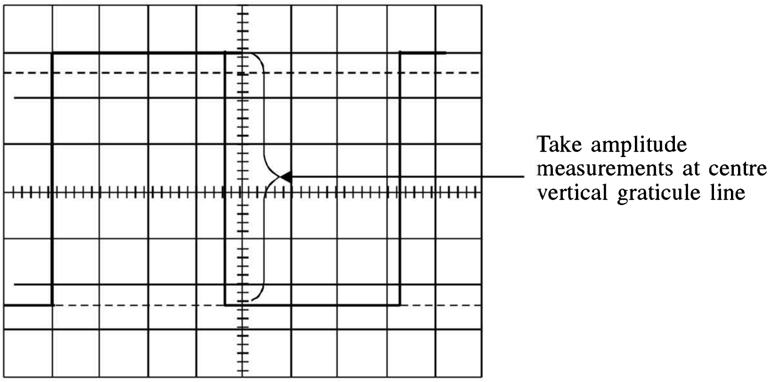

Take amplitude measurements at centre vertical graticule line

Figure 3.40 Voltage measurement on the centre vertical graticule line.

Many oscilloscopes have on-screen cursors that let us take waveform measurements automatically on-screen, without having to count graticule marks. (A cursor is simply a line that we can move across the screen.) Two horizontal cursor lines can be moved up and down to bracket a waveform's amplitude for voltage measurements, and two vertical lines move right and left for time measurements. A readout shows the voltage or time at their positions.

3.7.2 Time and Frequency Measurements

We take time measurements using the horizontal scale of the oscilloscope. Time measurements include measuring the period, pulse width, and timing of pulses. Frequency is obtained as the reciprocal of the period. Like voltage measurements, time measurements are more accurate when we adjust the portion of the signal to be measured to cover a large area of the screen. Taking time measurement along the centre horizontal graticule line, having smaller divisions, makes for the best time measurements (see Figure 3.41.)

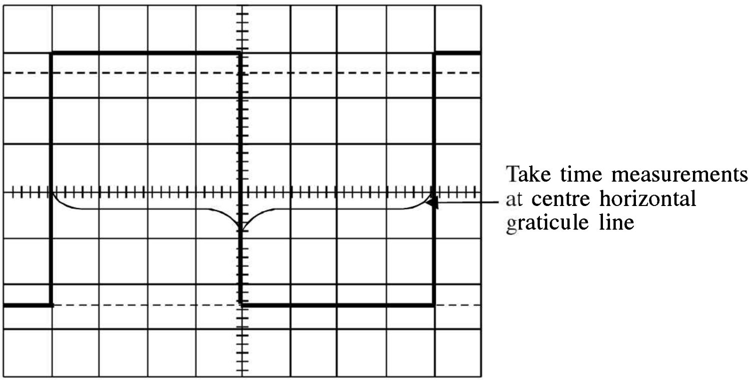


Figure 3.41 Time measurement on the centre horizontal graticule line.

3.7.3 Pulse and Rise Time Measurements

In many applications, the details of a pulse's shape are important. Pulses can become distorted and cause a digital circuit to malfunction, and the timing of pulses in a pulse train is often significant.

Standard pulse measurements are *pulse width* and *pulse rise time*. Rise time is the amount of time a pulse takes to go from the low to high voltage. By convention, the rise time is measured from 10% to 90% of the full voltage of the pulse. This eliminates any irregularities at the pulse's transition corners. This also explains why most oscilloscopes have 10% and 90% markings on their screen. Pulse width is the amount of time the pulse takes to go from low to high and back to low again. By convention, the pulse width is measured at 50% of full voltage.

Pulse measurements often require fine-tuning the triggering. To become an expert at capturing pulses, we should learn how to use trigger holdoff and how to set the digitizing oscilloscope to capture pretrigger data. Horizontal magnification is another useful feature for measuring pulses, since it allows us to see fine details of a fast pulse.

The horizontal control section may have an *XY* mode that lets us display an input signal rather than the time base on the horizontal axis. This mode of operation opens up a whole new area of phase shift measurement techniques.

The phase of a wave is the amount of time that passes from the beginning of a cycle to the beginning of the next cycle, measured in degrees. Phase shift describes the difference in timing between the two otherwise identical periodic signals.

One method for measuring phase shift is to use the *XY* mode. This involves inputting one signal into the vertical system as usual and then another signal into the horizontal system. (This method only works if both signals are sine waves.) This setup is called an *XY* measurement because both the *X*- and *Y*-axis are tracing voltages. The waveform resulting from this arrangement is called a Lissajous pattern (named after the French physicist Jules Antoine Lissajous). From the shape of the Lissajous pattern, we can find the phase difference between the two signals as well as their frequency ratio. Figure 3.42 shows the Lissajous patterns for various frequency ratios and phase shifts.

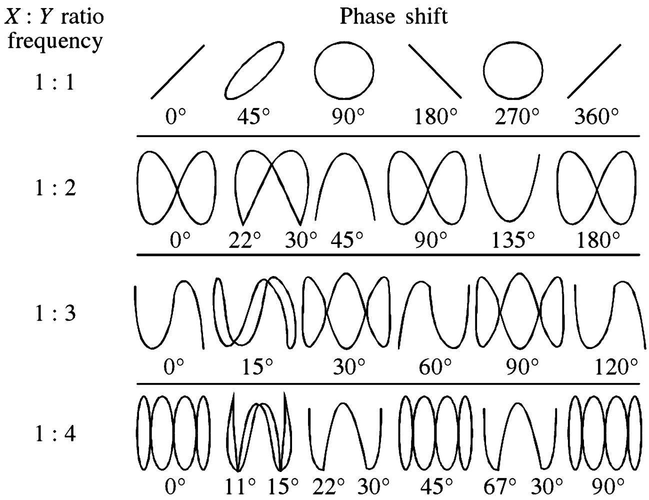


Figure 3.42 Lissajous patterns.

The *XY* measurement mode originated with analog oscilloscopes. Due to their relatively low sample density, DSOs may have difficulty creating real-time *XY* displays. Some DSOs create an *XY* image by accumulating data points over time, then displaying the composite. Digital Phosphor Oscilloscopes, on the other hand, are able to acquire and display a genuine *XY* mode image in real-time, using a continuous stream of digitized data. DPOs can also display an *XYZ* image with intensified areas.

3.7.4 Fully Automated Measurements with Digitizing Oscilloscopes

Digitizing oscilloscopes have functions that make waveform measurements easier. Modern DSOs and DPOs have front-panel buttons or screen-based menus from which we can select fully automated measurements. These include amplitude, period, rise/fall time, and much more. Many digitizing instruments also provide mean and rms calculations, duty cycle, and other math operations. Automated measurements appear as on-screen alphanumeric readouts. Typically, these readings are more accurate than those that are possible to obtain with direct graticule interpretation.

3.8 SELECTING AN OSCILLOSCOPE

The first step in selecting an oscilloscope is to know about the typical characteristics of the signals we need to capture and view. Are our signals of a complex nature? Are the signals repetitive or single shot? What is the bandwidth and/or rise time of the signals or transients we need to capture? On what signal characteristics do we need, to trigger—runts, pulse widths, slew rates, or glitches? How many signals do we need to display simultaneously? The answers to these questions will determine the type of scope we need. Some important parameters are given below.

3.8.1 The "5 Times" Rule for Bandwidth

An oscilloscope's bandwidth specification gives an idea of the instrument's ability to handle high-frequency signals. If the bandwidth of the scope is inadequate, high frequency information will be filtered out and critical timing information may be displayed inaccurately. It is important to remember that a scope's bandwidth is derived from how well it reproduces a sine wave. If we are dealing with high-speed digital electronics, we may never see a sine wave. A good rule of thumb for determining the bandwidth we need, is to use the "5 times rule"—to figure out the bandwidth of the scope we need, multiply the highest frequency component of the signal we want to examine by five. This will typically give better than 2% accuracy in the measurements.

In some applications we may not know the bandwidth of the signal of interest, but we may know the rise time. The frequency response of most scopes is designed so that a constant allows us to relate the bandwidth and rise time of the instrument using the equation

$$-3 \text{ dB bandwidth} \times \text{risetime} = K$$

where the –3 dB bandwidth is in hertz, rise time is the 10% to 90% rise time in seconds and K is a constant that depends on the oscilloscope's input filter type.

As a general rule of thumb, when selecting an oscilloscope we can assume that the bandwidth to rise time ratio is about 0.35. Using this rule to calculate bandwidth, will give us a figure that is sufficient for scope selection.

3.8.2 Sample Rate

When considering digitizing oscilloscope characteristics and specifications, it is important to make a clear distinction between sample rate and waveform capture rate. A digitizing oscilloscope's sample rate is specified as samples per second and indicates how frequently the scope samples the input signal. With higher sample rates, the signal is sampled more frequently allowing greater signal resolution.

The importance of adequate sample rate

The Nyquist sampling theorem states that the signal of interest must be sampled at least twice the rate of its highest frequency component to accurately reproduce it. While this is true in theory, sometimes it is not enough in practice. In reality, today's high-speed digital signals should be sampled at least at five times the rate of the highest frequency component (2.5 times the Nyquist rate) in order to be accurately reconstructed. Real-time sampling allows full bandwidth single-shot acquisition and eliminates the need for repetitive sampling techniques.

3.8.3 Triggering

Extended triggering capabilities offered in most oscilloscopes allow us to capture complex signals for device characterization or troubleshooting. They also provide more selective trigger control than is available with conventional edge triggering. In addition, their sophisticated user interfaces allow rapid setup of trigger parameters to increase productivity.

These oscilloscopes allow us to trigger on pulses defined by amplitude (such as runt pulses), qualified by time (pulse width, glitch, slew rate, setup-and-hold and time-out), and delineated by logic state or pattern (logic triggering). Combinations of extended and conventional triggers also help display video and other difficult-to-capture signals. Advanced triggering capabilities such as these deliver a large degree of flexibility when setting up testing procedures and can greatly simplify measurement tasks.

3.8.4 Record Length

Maximum oscilloscope record lengths are available from 1 K points to 8 M points. If we are simply monitoring an extremely stable sinusoidal signal, for example, a 500 point record length may be adequate. But if we are trying to find the causes of timing anomalies in a complex digital data stream, a long record length is more appropriate. In addition, long record lengths provide the detail we need to capture and analyze single-shot events.

3.9 ANALOG vs DSO vs DPO

Digital storage and digital phosphor oscilloscopes, though latest, still share bench-top space with aging analog oscilloscopes because both of these platforms have mutually exclusive strengths.

The DSO provides simultaneous multi-channel operations, as well as measurement automation, waveform storage, and I/O operations. The analog oscilloscope's intensity-graded display, with its varying brightness and fast waveform capture rate, innately brings a real-time *statistical* dimension to the viewed waveform. It indicates the frequency of occurrence of different parts of the signal. But the analog scope lacks the DSO's storage capabilities and other features. Consequently, it may be difficult to depend entirely upon either oscilloscope architecture for all the signal characterization needs.

DPO combines the best of the analog and digital worlds while going beyond both technologies. With one instrument, it is now possible to capture all of the salient information about a waveform in three dimensions—amplitude, time and the intensity axis that reveals amplitude distribution over time. The distribution of amplitude over time enables us to obtain an intensity-graded colour display that varies the displayed intensity levels with the signal's frequency of occurrence. The fast update rate of DPOs also makes it easier to capture and display infrequent waveforms or waveform variations.

Packetized signals made up of multiple components with relatively long time periods are especially difficult to capture faithfully with a DSO. One example is the composite video signal. It requires capturing a long time interval (and thus the use of a slow time-base setting) to understand the characteristics of the whole envelope. But in order to capture individual pulse detail, a fast time-base setting is required.

The normal procedure is to set the DSO's time base (and therefore its sample rate) to a horizontal rate slow enough to acquire the whole signal envelope. On a DSO, the slow sample rate produces aliasing on the individual faster pulses within the signal, a by-product of using a sample rate that is too low relative to the frequencies being measured. The result is a waveform that is distorted almost beyond recognition. Worse still, it can actually appear to be a lower-frequency waveform than it really is.

The solution, until now, has been to use an analog scope for viewing these types of signals. The analog display is regarded as a correct waveform profile. But it offers no way to store, automatically measure, or analyze the signal.

The DPO's abundance of waveform data, with 100 million points per second being sent to the display, solves the aliasing problem. The resulting waveform is clear and comprehensible, even though it was acquired at the slow time-base setting.

Further, because the DPO stores the waveform data in the dynamic three-dimensional database, statistical information about the signal can easily be derived. The DPO's internal histogram function gathers quantitative information on the distribution of the signal on-the-fly or on a stored waveform. The three-dimensional database can also be exported via the scope's GPIB port to an external PC for analysis including three-dimensional plots. This data presents a three-dimensional view wherein the frequency of occurrence shows up as the Z-axis of the graph. Like the DPO screen display itself, colour can be used to enhance appearance.

QUESTIONS

3.1 Draw a simple circuit diagram of a laboratory-type analog cathode ray oscilloscope (CRO) and describe its operation.

3.2 Explain random and raster scan. What is meant by 1024 × 768 resolution?

3.3 Explain the function of the following controls:

(a) Z modulation

(b) Astigmatism

(c) Delayed Sweep

(d) ALT/CHOP mode

3.4 Draw the block diagram of a dual trace digital storage oscilloscope.

3.5 Explain sampling methods used in digital oscilloscopes.

3.6 Explain the triggering controls used in analog and digital oscilloscopes. What are the special trigger settings available only in digital oscilloscopes?

3.7 Differentiate between analog oscilloscope, digital storage oscilloscope and digital phosphor oscilloscope.

3.8 List five parameters for each of the following subsystems of a digital storage oscilloscope.

(i) Time Base system
(ii) Vertical System
(iii) Triggering System
(iv) Measurement System
(v) Display System
(vi) Waveform Processing System

3.9 What are the important parameters used to select an oscilloscope?

3.10 What type of oscilloscope will you prefer for analysing video signals and why?

3.11 On what parameters does the deflection of the electron beam in an oscilloscope depend? Explain the effect of each parameter on the deflection.

3.12 Write typical specifications of a digital oscilloscope.

Chapter 4

Probes

4.1 INTRODUCTION

A probe can be any conductor used to establish a connection between the circuit under test and the measuring instrument. This conductor could be a piece of bare wire, a multimeter lead, or a piece of unterminated coaxial cable. These simple probes, however, do not fulfil the essential purpose of a probe, that is, to extract minimal energy from the circuit under test and transfer it to a measuring instrument with maximum fidelity. The bare wire can load the input amplifier with its high capacitance and inductance or even cause a short circuit; multimeter leads are unshielded and are often susceptible to stray pickup, and the unterminated coax will severely capacitively load the circuit under test (100 pF per metre typically). Also, the unterminated coaxial cable is usually resonant at certain frequencies and does not allow faithful transfer of the signal to the test instrument due to reflections. Emphasis should be on minimizing the reflections and other effects associated with unterminated coaxial probe cables and on reducing the effect of coaxial probe cables' capacitance in high-speed probing.

The ideal probe offers the following key attributes:

- Ease of connection
- Absolute signal fidelity
- Zero signal source loading
- Complete noise immunity

But in practice it is almost impossible to achieve these properties. In this chapter, we will study the various factors affecting the performance of a probe, the basic probe types used in various applications and look at some of the more common issues influencing probe selection.

4.2 TYPES OF PROBES

The types of probes which are available include general-purpose passive voltage probes, high voltage passive probes, 50 ohm divider passive voltage probes (Z_0, low impedance/high frequency), current probes, active (FET) voltage probes and differential voltage probes. Figure 4.1 shows the various categories of probes based on the signal to be measured.

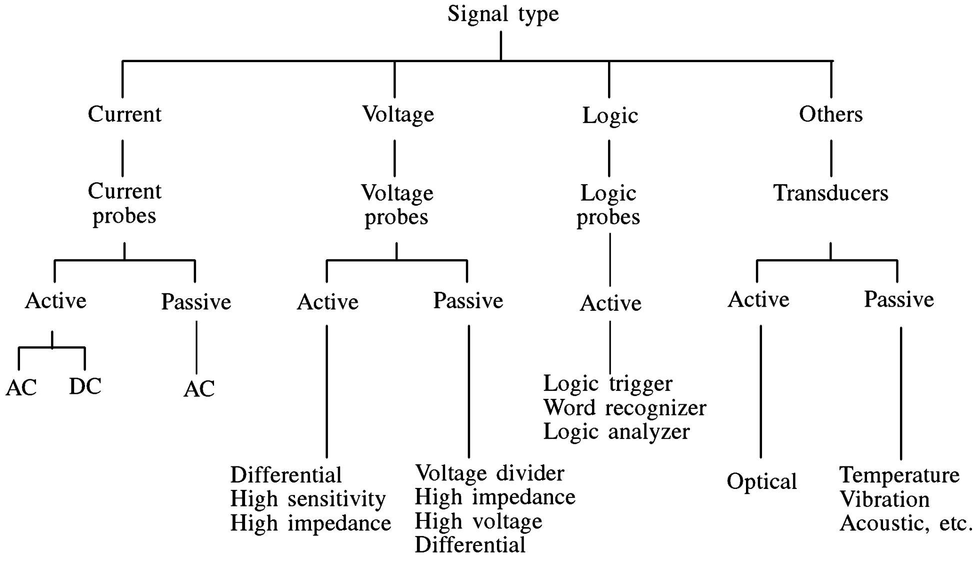


Figure 4.1 Various categories of probes.

Voltage signals are the most commonly encountered signal type in electronic measurements. Therefore, the voltage-sensing probe is the most common type of probe. Other types of probes are, in essence, transducers that convert a sensor output to a corresponding voltage signal. For example, the current probe transforms a current signal into a voltage signal, various transducers convert signals from optical, mechanical, thermal, acoustic, and other sources into voltage. When this is done, the transducer becomes the signal source for the purposes of selecting a probe.

4.2.1 Passive Voltage Probes

The passive voltage probes include low-impedance resistive divider probes and compensated high-resistance, passive divider probes. A subset of the compensated high-resistance passive divider probes is the high-voltage probe. This probe is required for very large signals.

Low-impedance resistive divider probe

Figure 4.2 shows the low-impedance resistive divider probe. The tip contains a resistor R_1.

The probe cable with a resistance of R_C, is terminated with the input resistance R_2 of the scope. The resistances R_1 and R_C are chosen depending upon the value of R_2 and the attenuation required.

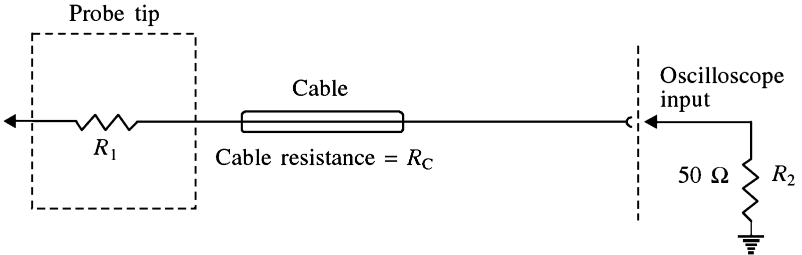


Figure 4.2 Low-impedance resistive divider probe.

Suppose that the input resistance of an oscilloscope is 50 Ω and we need attenuation factors of 1, 10 and 20. We choose a cable with resistance of 50 Ω and R_1 = 0, 450, and 950 Ω. They give 1:1, 10:1 and 20:1 attenuation, respectively. The probes are named 1X, 10X and 20X passive voltage probes. These are also called 50 ohm divider (Z_0) probes because the cable resistance matches with the input resistance of the oscilloscope. These probes provide the lowest input capacitance (typically < 1 pF for high frequency signals) and are used with high frequency, 50 ohm input scopes. They have the following features:

- Low capacitive loading and a very high bandwidth. They exhibit a frequency response that is essentially flat throughout their designed frequency range. As a result, we get very accurate timing measurements.
- Relatively low cost compared to the active probe.
- Useful for probing low-voltage signals (less than 50 V), such as ECL circuits and 50 Ω transmission lines.
- Relatively heavy resistive loading.

High resistance passive divider probe

This is a commonly used probe (see Figure 4.3) to measure low-to-intermediate signals (less than 500 V). Typically values are R_1 = 900 kΩ, R_2 = 1 MΩ, and R_3 = 111 kΩ. The cable is a high-impedance cable. A termination box at the end of the probe cable connects to the input of the oscilloscope; R_3 is contained within the termination box that works in parallel with the input resistance of the oscilloscope. This results in a 10:1 attenuation. The need for the capacitors C_1 and C_3 will be discussed later.

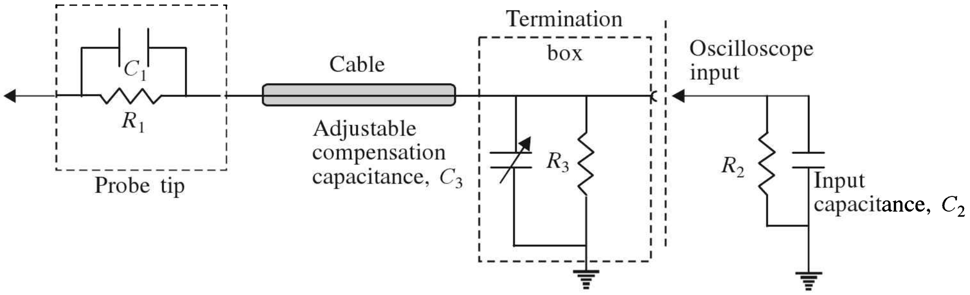


Figure 4.3 A high resistance passive divider probe.

High resistance passive divider probes have the following features:

- High resistance (1 MΩ or greater)
- Adjustable compensating capacitor for matching to the input capacitance of the oscilloscope
- High dynamic range
- Most rugged
- Low cost compared to active probes
- Lowest bandwidth
- Heaviest capacitive loading compared to the low-impedance resistive divider probe. So, there is a decreasing impedance level with increasing frequency.

Figure 4.4 shows a typical passive probe with accessories.

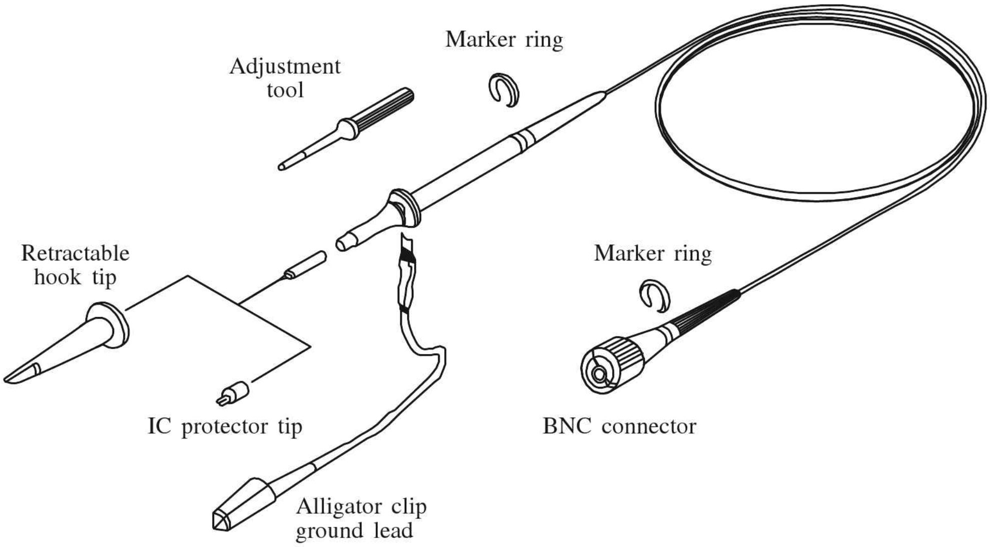


Figure 4.4 A typical passive probe with accessories.

High voltage passive probe

This is basically a high resistance passive divider probe with high probe tip resistance (typical value is 500 MΩ). So the attenuation factor is very high. Several high voltage probes are available that provide 100X or 1000X compensated dividers. Because of the larger attenuation factors required for high voltage applications, the input capacitance is reduced (to approximately 3 pF). These probes are used for input voltage > 500 V. They have the following features:

- Very high dynamic range. Therefore, very high voltages can be measured with this probe, and it is the safest at voltages over 500 V. Typical oscilloscope probes could break down if they are used at such high voltages
- Physically larger in size
- Lower bandwidth compared to standard oscilloscope probes.

4.2.2 Active Voltage Probes

Active probes contain active components, such as transistors, for their operation in addition to the *RC* network. Most often, the active device is a field effect transistor (FET). The advantage of a FET input is that it provides a high input resistance (typically >100 kΩ) and very low input capacitance (typically < 0.4 pF to < 2 pF). Such ultra-low capacitance has several desirable effects. Figure 4.5 shows an active probe.

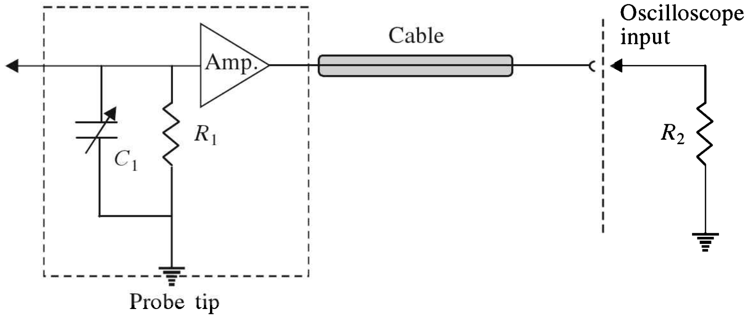


Figure 4.5 An active probe.

Some FET probes include an offset control that allows a substantial increase in the active probe's linear dynamic range. These probes are very commonly used for 50 Ω load applications. FET probes, then, have a 50 Ω output impedance and can drive a 50 Ω cable. This capability allows the distance from the probe tip to the instrument to be increased within the practical limits of the probe amplifier system and the limitations of the coaxial cable. Features include:

- Very high input resistance provides minimum loading
- Very low input capacitance provides high input impedance over a broad band of frequencies
- Higher bandwidth (up to about 5 GHz)
- Highest cost of all the probes
- Limited dynamic range. This is usually ±40 V dc plus peak ac.

Differential probes

Figure 4.6 shows a differential probe. The significant difference here is that the probe tip has two inputs, a positive (or noninverting) and a negative (or inverting) input. These two inputs feed a differential amplifier, which in turn is connected to the input of the scope through a matching cable. For a 50 Ω input scope, amplifier drives a 50 Ω cable that is connected to the input of the scope. One benefit of the differential probe, compared to subtracting two channels with passive probes, is that the electrical path for the two signals is matched as closely as possible. This provides substantially higher CMRR performance over a broader frequency range. A bandwidth of 1 GHz can be achieved with CMRR performance ranging from 60 dB (1000:1) at 1 MHz to 30 dB (32:1) at 1 GHz. A true differential probe allows the

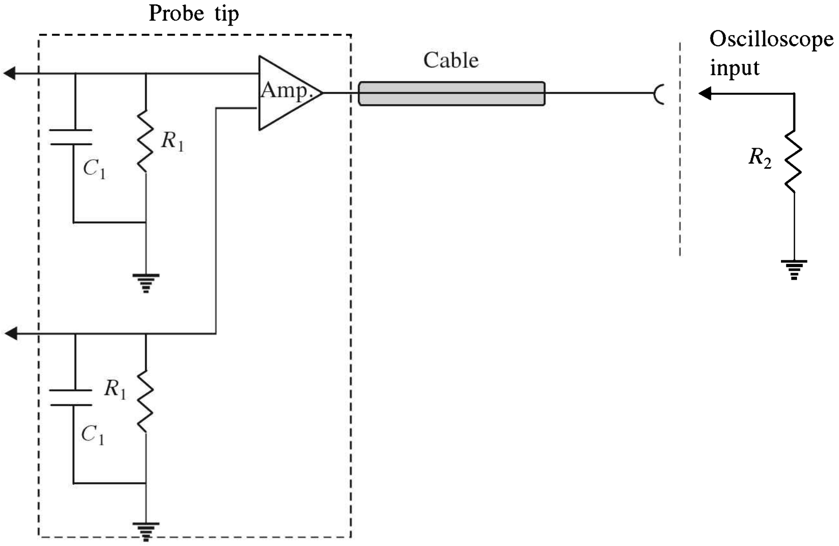


Figure 4.6 Differential probe.

user to adjust the probe's attenuation for compatibility with a variety of amplifiers to provide higher common mode rejection ratio. This attenuation adjustment includes probe compensation so that the probes match at high frequencies as well as low frequencies. Differential probes have the following features:

- Controls for dc offset, dc reject, and coupling are available.
- High common mode rejection. Note that the higher the common mode rejection ratio the smaller the signals that are possible to view. The benefits are ease of viewing small signals in the presence of large dc offsets or other common mode signals, and the accuracy that can be achieved in looking at these signals.
- The probe is more bulky.
- A differential probe is more expensive than using two passive probes.
- Less dynamic range.
- Requires the use of an external power and control module.

4.2.3 Current Probes

Current probes provide a method to measure the current flowing in a circuit. Two types of current probes are available—the traditional ac only probes and the *Hall Effect* semiconductor ac/dc probe.

AC current probe

Alternating currents induce voltage in a transformer resulting from the buildup and collapse of flux fields as the current changes direction. This transformer action is the basis for ac current probes. The ac current probe head is actually a coil (primary) that has been wound to precise specifications on a magnetic core. When this probe head is held within a specified orientation and proximity to an ac current carrying conductor, voltage on the secondary coil appears. This probe voltage is directly proportional to ac current in the primary coil. This current-related voltage can be displayed as a current-scaled waveform on a scope.

The probe has a set of jaws that encloses the wire that the measured current is flowing through. Measured current wire acts as the primary of the transformer. No electrical connection is needed. The circuit does not have to be broken or altered in any way, as the current probe measures whatever current is passing through its closed jaws. If the sensitivity of the probe and oscilloscope combination is too low for a particular measurement to be made, several turns of the current carrying wire can be inserted into the jaws. The probe will effectively have a larger current to measure (the original current times the number of turns). In a similar manner, the difference between two currents can be measured if the two wires in question are inserted with the currents flowing in opposite direction (the sum will be measured if the currents are flowing in the same direction). The physical size of the wires and the current probe will be a factor in determining how many wires can be inserted. Although the probe does not require a direct electrical connection, it still removes energy from the circuit under test. Normally, this small amount of energy loss will not disturb the circuit, but can be a factor in some cases.

The bandwidth for ac current probes depends on the design of the probe's coil and other factors. Bandwidths as high as 1 GHz are possible. However, bandwidths under 100 MHz are more typical.

The amp-second product defines the maximum limit for linear operation of any current probe. This product is defined for current pulses as the average current amplitude multiplied by the pulse width. When the amp-second product is exceeded, the core material of the probe's coil goes into saturation. Since a saturated core cannot handle any more current-induced flux, there can no longer be constant proportionality between current input and voltage output. The result is that the waveform peaks are *clipped off* in areas where the amp-second product is exceeded.

Alternating current probes are passive in that they do not require external power. A steady-state direct current, however, will not induce a current in a transformer. So this probe cannot be used to measure direct currents.

AC/DC current probe

The other type of current probe works using the principle of the Hall effect. The Hall effect produces voltage output when direct current is applied to a semiconductor device present in dc generated magnetic field. This technique requires the use of an external power supply, but does measure both alternating and direct current. Consequently, a direct current probe is an active device, in that it requires external power.

A dc probe amplifier may also be required for combining and scaling the ac and dc levels to provide a single output waveform for viewing on an oscilloscope.

Each type of current sensing device will roll off or produce a nonlinear output at a certain point. A given ac probe will roll off at a certain maximum current amplitude and frequency, as well as at a minimum amplitude and low-end frequency. DC probes will be linear from dc to a maximum frequency and will have a minimum and maximum sensitivity level as well. Since we often encounter signals which contain both ac and dc components, it is important to be able to measure both components simultaneously with a single probe. In addition, signals which may look like ac, such as a square wave switching on and off, but never fall below zero amps, actually have a significant dc component and cannot be accurately captured with a transformer only ac probe. A unique solution is to combine both a Hall effect device and a transformer to provide broadband current measurement capabilities within one system.

AC only current probes have frequency response from a few hundred Hz to GHz. Combining a Hall effect device with an ac transformer provides a frequency response from dc to 50 MHz. A current probe typically imposes less loading than that imposed by other probes.

4.2.4 Logic Probes

These probes are used to detect faults in the digital systems. A logic analyzer is the primary tool for identifying and isolating fault occurrences by displaying simultaneously data on many channels. Many times, the actual cause of the logic fault is often due to the analog attributes of the digital waveform. These faults may be pulse width jitter, pulse amplitude *aberrations*, analog noise, and crosstalk. To analyze the analog attributes of digital waveforms requires the use of an oscilloscope. However, to isolate exact causes, designers look at specific data pulses occurring during specific logic conditions. This requires a logic triggering capability that is more typical of a logic analyzer than an oscilloscope. Such logic triggering can be added to most oscilloscopes through use of a word recognizer trigger probe.

The logic probe can accept a number of digital signals through data channel probes which terminate at one end of the probe head. The other end of the probe head is connected to the oscilloscope's external trigger input. The logic probe head contains miniature switches and digital comparators. The switches are programmed for the desired sequence of the data input. When the data input matches the programmed sequence, the probe head generates a trigger pulse for the oscilloscope. At this instance a particular digital signal waveform, which is fed to the vertical inputs of the scope, can be analyzed for any fault.

4.2.5 Optical Probes

During the development and verification of optical components, there is a need for general-purpose optical waveform measurement and analysis with the help of an oscilloscope. Optical probes are used for this purpose.

An optical probe is an optical-to-electrical converter. On the optical side, the probe is selected to match the specific optical connector and fiber type or optical mode of the device that is being measured. On the electrical side, the standard probe-to-scope matching criteria are adopted.

4.2.6 Other Probes

In addition to all of the above standard probe types, a variety of other special types of probes and probing systems are used. These include:

- Environmental probes, which are designed to operate over a wide temperature range, use appropriate sensors to convert the environmental variable into voltage which is measured by an oscilloscope.
- Temperature probes are used to measure the temperature of components and other heat generating items. They use temperature sensors like thermocouple, RTDs, etc. to convert temperature into voltage which is measured by an oscilloscope.
- Probing stations and articulated arms are used for probing fine-pitch devices such as multichip modules, hybrid circuits, and ICs. These are miniature probes that provide contact on the pins of small geometry devices.
- Logic analyzer probes are used to interface target systems which are to be analyzed. Because of a large number of input channels, it is not practical to use unique cables to connect individual channel. Instead eight or more channels are grouped into *pods* (Plug-on-devices). Each pod connects to the system as a single cable. Passive probes are most common. Typically, each channel is terminated with 100 kΩ and 8 pF. These probes being of small size can be terminated right at the point of connection to the target system. This avoids additional stray capacitance due to the wires from large active pods to the circuit under test.

Many a times, logic analyzer probes have a probe adapter or a preprocessor. A preprocessor provides not only fast, reliable and correct mechanical connection to the target system but also necessary electrical adaptation like clocking and demultiplexing to capture system's operation correctly. Some microprocessors prefetch information from memory that may never get executed. Preprocessors can also distinguish prefetched information from executed. Preprocessors also come with inverse assembler to decode HEX information to microprocessor mnemonics.

4.3 PROBE LOADING AND MEASUREMENT EFFECTS

Ideally, a probe should simply be a non-intrusive (infinite resistive, non-capacitive and non-inductive) wire attached between oscilloscope and the circuit of interest so that it provides an exact replica of the signal being probed. But in practice, the probe becomes part of the circuit and introduces resistive, capacitive and inductive loading to the circuit. This in turn causes the scope to provide deviations in the measurement depending upon the loading of the probe. Since it is impossible to totally eliminate loading, the goal should be to choose a probe that will minimize the loading effect and provide the most accurate replication of the signal.

4.3.1 Resistive Loading

Resistive loading is significant if the input resistance of the probe is of the same magnitude as of the signal being probed. When a probe is introduced into a circuit, some of the current that had been flowing in the circuit will now flow into the probe, reducing the voltage at the point being probed. Figure 4.7 shows a circuit to understand the effects of resistive loading. V_S is

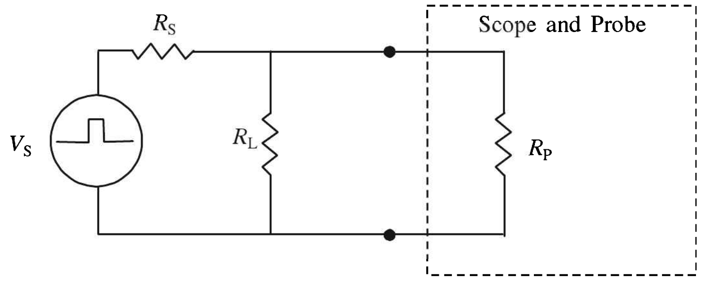


Figure 4.7 Resistive loading.

the source voltage, R_S the source resistance, R_L the load resistance and R_P is the equivalent resistance of the scope and the probe which is connected across the load.

Let us assume that all resistors have equal value. When the probe is not connected across the load, voltage across $R_L = V_S/2$. However, when the probe is connected, effective load resistance is $R_L \| R_P$. This reduces the voltage across R_L to $V_S/3$, which causes more than 30% error in measurement due to loading of the probe. Therefore, the probe resistance should be greater than 10 times the resistance of the circuit under test, so that the amplitude measurement errors are less than 10%. Alternatively, the effective circuit resistance seen by the probe should be as low as possible.

Thus, resistive loading results in attenuation of the amplitude, dc offset shift, and circuit bias change. It may also cause a malfunctioning circuit to start working or more likely cause a functioning circuit to become nonfunctional.

4.3.2 Capacitive Loading

Probes also have some capacitance that affects timing measurements. This capacitance introduces measurement errors that are frequency dependent. As signal frequencies or transition speeds increase, the capacitive load becomes predominant. Consequently, capacitive loading becomes a matter of increasing concern. In particular, capacitive loading affects the rise and fall times on fast-transition waveforms and also the amplitudes of high-frequency components in waveforms.

Effect on rise time

Figure 4.8 shows the equivalent circuit of the probe and the source and the capacitive loading effect on rise time. R_S and C_S are equivalent source resistance and source capacitance and R_P and C_P are probe resistance and capacitance.

We can see from Figure 4.8(b) that the rise time (time taken for the waveform to rise from 10% level to 90% level) with the probe, t_{rp}, is higher than the rise time without the probe t_r. This loading is caused by the tendency of the probe's capacitance to act as a low-pass filter at high frequencies. It shunts the high-frequency information to ground and significantly reduces the probe's input impedance at high frequencies. The rise time can be represented in terms of component values of the circuit and is given by the equation

$$\text{Rise time} = 2.2RC$$

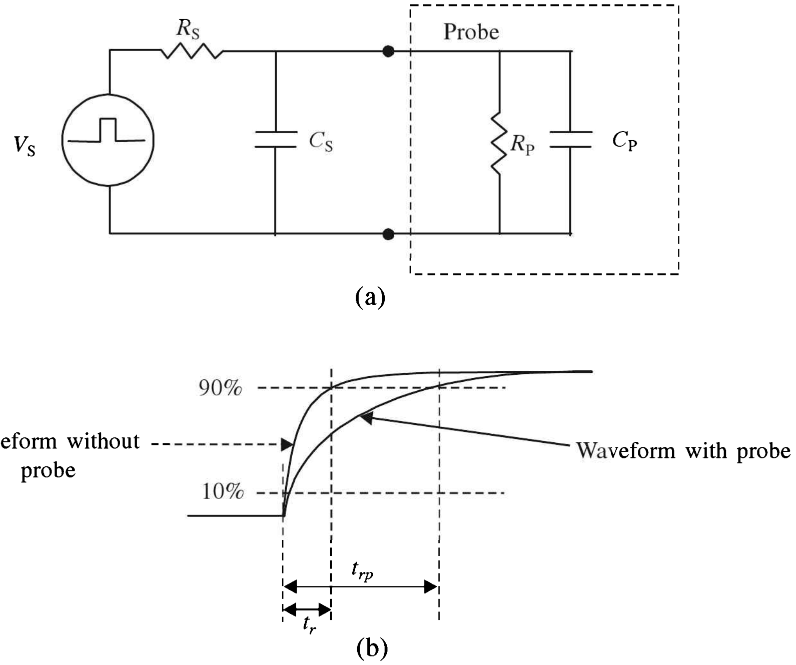


Figure 4.8 Capacitive loading.

where $R = R_S||R_P$ and $C = C_S + C_P$.

We can estimate the effect of probe capacitance on rise time by taking the ratio of the probe's specified capacitance to the source capacitance. For example, let $R_S = 50\ \Omega$, $C_S = 20$ pF, $R_P = 1$ MΩ (>>R_S) and C_P =10 pF. The estimate of percentage change in rise time is determined below:

$$\frac{C_P}{C_S}\times 100\% = \frac{10}{20}\times 100\% = 50\%$$

For passive probes, the greater the attenuation ratio, the lower the probe capacitance in general. Where smaller probe capacitance is needed, active FET input probes should be used. Active probes with probe capacitance of 1 pF and less are available.

Effect on impedance

At high frequency the reactance of the capacitance decreases. This reduces the impedance offered by the probe, resulting in the voltage drop across the probe. So, not only is the rise time affected, but also the amplitude reduces with capacitive loading.

Further, since impedance is a complex quantity (involves both magnitude and phase), change in reactance of the capacitance also affects the phase angle. Figure 4.9 shows the variation of magnitude and phase shift of the impedance of a typical passive probe versus frequency.

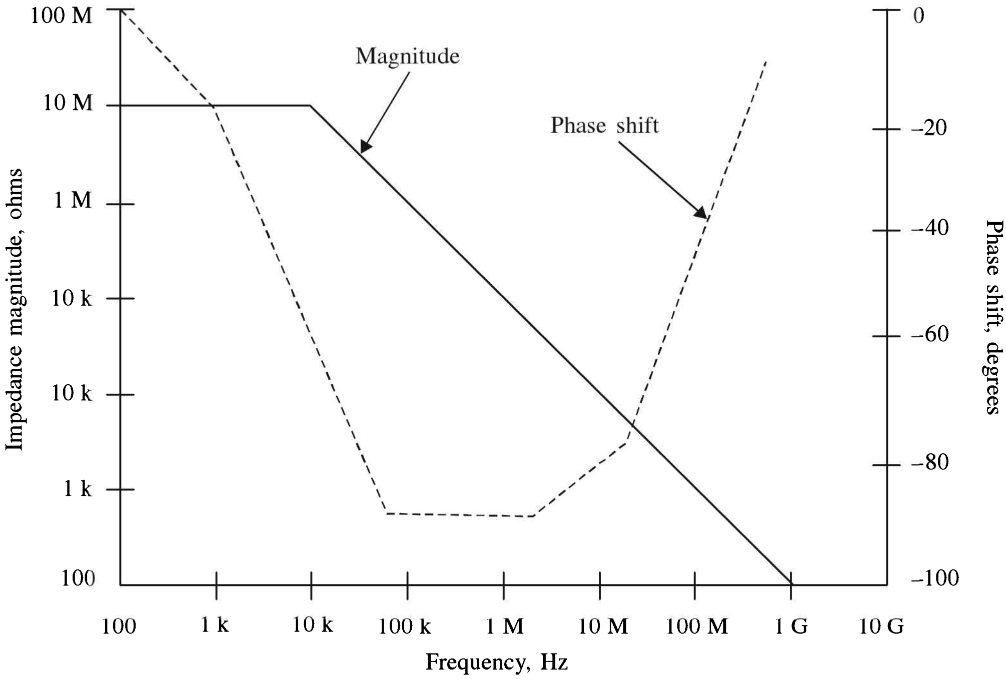


Figure 4.9 Impedance variation versus frequency of a typical passive probe

Figure 4.10 shows the impedance variation versus frequency of typically different probes.

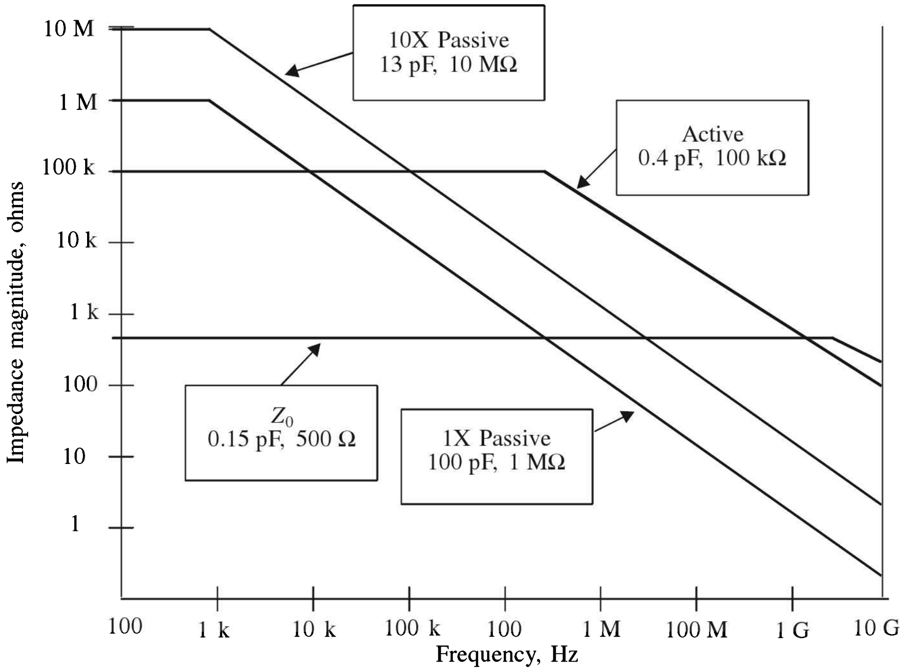


Figure 4.10 Impedance variation versus frequency of various typical probes.

Effect on bandwidth

Bandwidth, which is defined as the point on an amplitude versus frequency plot where the amplitude is 3 dB down from the reference level, is also affected. When the rise time is slowed, there is a corresponding decrease in the bandwidth. This decrease comes from the equation:

$$\text{Bandwidth} \approx \frac{0.35}{\text{rise time}}$$

When the probe is connected to the oscilloscope, bandwidths of both the probe and the oscilloscope are considered. The oscilloscope's bandwidth should exceed the predominate frequencies of the signals to be measured, and the bandwidth of the probe used should be equal to or should exceed the bandwidth of the scope. The real concern is the bandwidth at the probe tip. Sometimes, manufacturers specify bandwidth at the probe tip for certain oscilloscope/probe combinations.

A test setup for verifying bandwidth at the probe tip is given in Figure 4.11. The test signal source is specified to be a 50 Ω source terminated in 50 Ω, resulting in an equivalent 25 Ω source termination. Additionally, the probe is connected to the source by a probe-tip-to-BNC adaptor or its equivalent. This requirement for probe connection ensures the shortest possible ground path. Using this test setup, a 100 MHz scope/probe combination results in an observed rise time of ≤ 3.5 ns. ($t_r \approx 0.35/\text{BW}$).

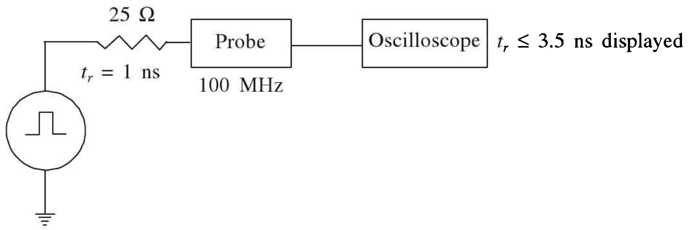


Figure 4.11 Setup to verify probe tip bandwidth.

Effect on probe attenuation ratio

Probe attenuation ratio is also an important consideration. The oscilloscope must have enough gain to allow viewing of the attenuated signal when using probes with larger attenuation ratios. When an attenuating probe is used with an oscilloscope, the input resistance and input capacitance of the oscilloscope represented by R_2, C_2 and the probe resistance and capacitance represented by R_1, C_1 should satisfy the following equation for optimum signal transfer:

$$R_2C_2 = R_1C_1$$

To test if the probe is properly compensated, a low frequency square wave (1 to 10 kHz normally) is applied to the probe tip and the waveform is seen on the oscilloscope.

Figure 4.12 shows the types of displays which can appear on the scope.

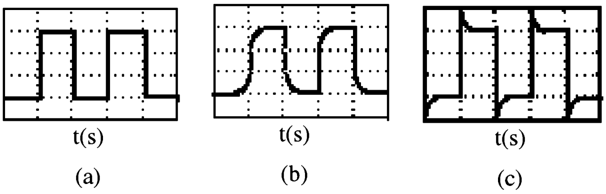


Figure 4.12 Probe compensation: (a) correctly compensated, (b) under-compensated, and (c) over-compensated.

Improper compensation results in either overshoot, roll-off or incorrect signal amplitudes. For changing the probe capacitance to attain the compensation, a variable capacitor is used in the probe tip or probe termination box.

The square wave is a wide bandwidth signal rich in harmonics. If the probe is properly compensated, it measures wide bandwidth signals with a minimum of waveform distortion. This concept is also used for square wave testing of amplifiers.

4.3.3 Inductive Loading

The inductive loading comes from the inductance of the probe ground lead. Basically, inductive loading distorts the signal being measured. Figure 4.13 illustrates the adverse effects of inductive loading on the circuit under test. The inductance of a typical ground path is 1 nH/mm. Thus, a typical probe ground lead of 7.5 cm has an inductance of 75 nH. The ringing that is introduced due to the loading has a frequency, f, equal to:

$$f = \frac{1}{2\pi\sqrt{LC}}$$

where C is the sum of the circuit capacitance and the probe capacitance and L is the ground lead inductance. As the inductance decreases, the ring frequency increases. The ringing amplitude can introduce about 50% amplitude error. The easiest way to reduce the inductive loading for any measured waveform is to use a shorter probe ground lead.

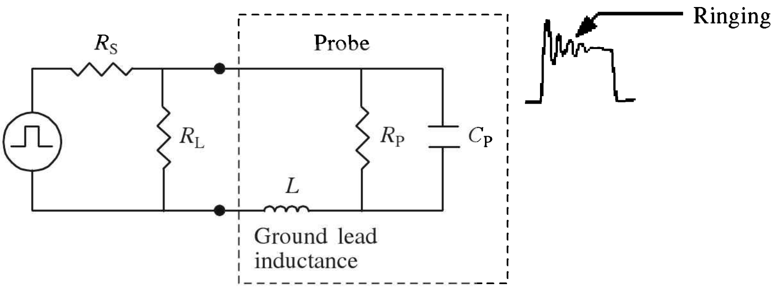


Figure 4.13 Ground lead inductance effect.

If the source ground lead is also long, it introduces another inductance in series with the probe ground lead inductance. The total inductance then is the sum of two inductances.

4.4 PROBE SPECIFICATIONS

Proper probe selection extends and enhances an instrument's performance, while wrong probe selection often reduces the system's performance. Understanding probe specifications helps to ensure that the performance of the instrument meets the application's requirements. Some of the specifications have already been studied in previous sections. These are summarized below. The specifications which are applicable to all the probes are marked *universal* and those applicable to a specific probe are marked accordingly.

Aberration (universal)

An aberration is any amplitude deviation from the expected or ideal response to an input signal. In practice, aberrations usually occur immediately after fast waveform transitions and appear as ringing. Aberrations are measured, or specified, as a ± percentage deviation from the final pulse response level (see Figure 4.14). This specification might also include a time window for the aberrations. For example: Aberrations should not exceed ±3% or 5% peak-to-peak within the first 30 nanoseconds.

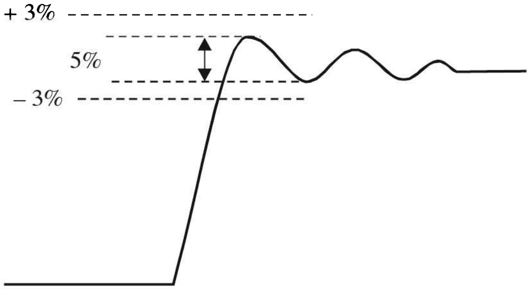


Figure 4.14 Defining aberration.

Attenuation factor (universal)

The attenuation factor is the amount by which the probe reduces signal amplitude. All probes have an attenuation factor, and some probes may have selectable attenuation factors. Typical attenuation factors are 1X, 10X, and 100X. A 1X probe does not reduce, or attenuate the signal, while a 10X probe reduces the signal to 1/10th of its probe tip amplitude. Probe attenuation factors allow the measurement range of an oscilloscope to be extended. For example, a 100X probe allows signals of 100 times greater amplitude to be measured.

Accuracy (universal)

For voltage-sensing probes, accuracy generally refers to the probe's attenuation of a dc signal.

The calculations and measurements of probe accuracy generally include the oscilloscope's input resistance. Thus, a probe's accuracy specification is only correct or applicable when the probe is being used with an oscilloscope having the assumed input resistance. An example of accuracy specification is: 10X within 3% (for scope input of 1 M Ω ±2%).

For current-sensing probes, the accuracy specification depends on the accuracy of the current-to-voltage conversion. This depends on the current transformer turns ratio and the value and accuracy of the terminating resistance. Current probes that work with dedicated amplifiers have outputs that are calibrated directly in amps/div and have accuracy specifications that are given in terms of attenuator accuracy as a percentage of the current/division setting.

Bandwidth (universal)

All probes have bandwidth. A 10 MHz probe has a 10 MHz bandwidth. The bandwidth of a probe is that frequency where the probe's response causes output amplitude to fall to 70.7% (–3 dB).

Some probes have a low-frequency bandwidth limit as well, for example ac current probes. Because of their design, ac current probes cannot pass dc or low-frequency signals. Thus, their bandwidth is specified with two values, one for low frequency and one for high frequency.

Capacitance (universal)

Generally, probe capacitance specifications refer to the capacitance at the probe tip. This is the capacitance that the probe adds to the circuit under test. Probe tip capacitance is important because it affects how pulses are measured. A low tip capacitance minimizes errors in making rise time measurements. Also, if pulse duration is less than five times the probe's *RC* time constant, the amplitude of the pulse is affected. Probes also present a capacitance to the input of the oscilloscope, and this probe capacitance should match that of the oscilloscope. This capacitance is referred to as a compensation range, which is different from tip capacitance. For probe matching, the oscilloscope's input capacitance should be within the compensation range of the probe.

CMRR (differential probes)

Common Mode Rejection Ratio (CMRR) is a differential probe's ability to reject any signal that is common to both test points in a differential measurement. CMRR is a key figure of merit for differential probes and amplifiers, and is defined as

$$\text{CMRR} = \left|\frac{A_d}{A_c}\right|$$

where A_d is the voltage gain for the difference mode signal, and A_c is the voltage gain for common-mode signal.

Ideally, A_d should be large, while A_c should be closer to zero, resulting in a CMRR approaching infinity. In practice, a CMRR of 10,000:1 (80 dB) is considered quite good.

This means a common-mode input signal of 5 V will be rejected to the point where it appears as 0.5 mV at the output. Such rejection is important for measuring difference signals in the presence of noise.

Since CMRR decreases with increasing frequency, the frequency at which CMRR is specified is as important as the CMRR value. A differential probe with a high CMRR at a high frequency is better than a differential probe with the same CMRR at a lower frequency.

Continuous wave (CW) frequency current derating (current probes)

Current probe specifications include amplitude versus frequency derating curves that relate core saturation to increasing frequency. The core saturation with increasing frequency means a waveform with an average current of zero amperes will experience clipping of amplitude peaks as the waveform's frequency or amplitude is increased. Wave shape can be sinusoidal or square.

Decay time constant (current probes)

The decay time constant specification indicates a current probe's pulse supporting capability. This time constant is the secondary inductance (probe coil) divided by the terminating resistance. The decay time constant is sometimes called the probe *L/R* ratio.

With larger *L/R* ratios, longer current pulses can be represented without significant decay or droop in amplitude. With smaller *L/R* ratios, long duration pulses will be seen as decaying to zero before the pulse is actually completed.

Direct current (current probes)

Direct current decreases the permeability of a current probe's coil core. This decreased permeability results in a decreased coil inductance and *L/R* time constant. The result is reduced performance for low frequencies and loss of measurement response for low frequency currents. Some ac current probes offer current bucking options that nullify the effects of dc.

Insertion impedance (current probes)

Insertion impedance is the impedance that is transformed from the current probe's coil (the secondary) into the current carrying conductor (the primary) that is being measured. Typically, a current probe's reflected impedance values are in the range of milliohms and present an insignificant effect on circuits of 25 Ω or more impedance.

Input capacitance (universal)

The probe capacitance measured at the probe tip is called the input capacitance.

Input resistance (universal)

A probe's input resistance is the impedance that the probe places on the test point at dc.

Maximum input current rating (current probes)

The maximum input current rating is the total current (dc plus peak ac) that the probe will accept and still perform as specified. In ac current measurements, peak-to-peak values must be derated versus frequency to calculate the maximum total input current.

Maximum peak pulse current rating (current probes)

This rating should not be exceeded. It takes into account core saturation and development of potentially damaging secondary voltages. The maximum peak pulse current rating is usually stated as an amp-second product.

Maximum voltage rating (universal)

Voltages approaching a probe's maximum rating should be avoided. The maximum voltage rating is determined by the breakdown voltage rating of the probe body or the probe components at the measuring point.

Probe coding (universal)

Probe coding provides the user with an indication of the actual probe tip sensitivity. This coding eliminates the need to divide by the attenuation ratio or remember which probe is being used. Most passive probes today provide readout capability that allows us to read our measurements directly from the oscilloscope screen.

Propagation delay (universal)

Every probe offers some small amount of time delay or phase shift that varies with signal frequency. This is a function of the probe components and the time it takes for the signal to travel through these components from probe tip to oscilloscope connector. Usually, the major shift is caused by the probe cable. For example, a 100 cm section of special probe cable has a 5 ns signal delay. For a 1 MHz signal, the 5 ns delay results is a two-degree phase shift. A longer cable results in correspondingly longer signal delays.

Rise time (universal)

A probe's 10% to 90% response to a step function indicates the fastest transition that the probe can transmit from tip to scope input. For accurate rise- and fall-time measurements on pulses, the measurement system's rise-time (scope and probe combined) should be three to five times faster than the fastest transition to be measured.

Temperature range (universal)

Current probes have a maximum operating temperature that is the result of heating effects from energy induced into the coil's magnetic shielding. Increasing temperature corresponds to increased losses. Because of this, current probes have a maximum amplitude versus frequency derating curve. Attenuator voltage probes (i.e. 10X, 100X, etc.) may be subject to accuracy changes due to changes in temperature.

Voltage derating with frequency (universal)

This specification is applicable to all high frequency probes. Either the termination elements or the resistive centre conductor in the probe cable limits the maximum voltage that may be applied to a probe at a specific frequency. Figure 4.15 shows a typical derating curve.

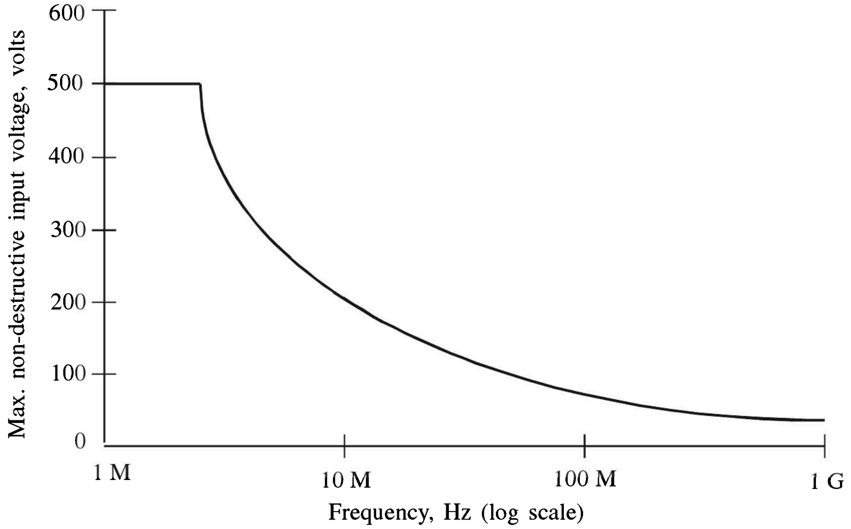


Figure 4.15 Voltage derating curve.

QUESTIONS

4.1 What is a probe? What are its key attributes? List the various types of probes with their applications.

4.2 Explain the design of passive and active probes with their features.

4.3 Explain the construction and features of ac and ac/dc current probes. Give at least four specifications especially applicable for current probes.

4.4 Explain resistive, capacitive, and inductive loading effects of probes.

4.5 A high impedance probe with 9 MΩ resistance and 4 pF capacitance is connected to an oscilloscope with an input resistance of 1 MΩ. When the probe was connected, the effective capacitance decreased to 3.6 pF. Find the capacitance of the oscilloscope alone.

4.6 How do you test whether the probe is properly compensated?

4.7 Figure Q4.7 shows the equivalent circuit of a probe with the ground lead inductance and source inductance. If L_{source} = 0.01 μH and $L_{\text{ground lead}}$ = 0.075 μH, find the frequency of ringing. If the actual bandwidth of the input pulse is 500 MHz, how will the ringing affect the signal at the end of the probe. Now if length of the cable is reduced so that $L_{\text{ground lead}}$ = 0.005 μH, what will be the ringing frequency and how it will affect the signal at the end of the probe?

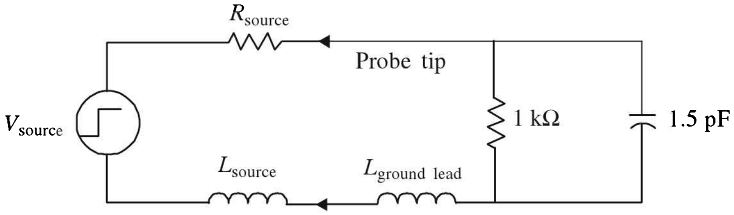


Figure Q4.7

4.8 Figure Q4.8 shows the bandwidth derating curve of a CRO. The horizontal scale in this figure shows the derating factor necessary to obtain amplitude accuracies better than 30%, i.e. 3 dB down point. With no derating (a factor of 1.0), a 100 MHz scope will have up to a 30% amplitude error at 100 MHz. If we want amplitude measurements to be within 3%, the bandwidth of this scope must be derated by a factor of 0.3 to 30 MHz. Anything beyond 30 MHz will have an amplitude error in excess of 3%. If this CRO is coupled with a probe of 100 MHz bandwidth, what should be the signal bandwidth so that the amplitude measurement is made within 3% error?

4.9 Show that the rise time of the capacitive loading of a probe is equal to 2.2 *RC* where *R* and *C* are the equivalent values of resistance and capacitance of the equivalent circuit.

4.10 A 1 V peak-to-peak 1 MHz sine wave is connected to both inputs of a differential probe (with gain equal to 1) and the probe output is 0.33 mV peak-to-peak. Now the signal is fed differentially to the probe. Signal amplitude is adjusted so that the output of the probe is again 0.33 mV peak-to-peak. The input to the probe is now measured, also as 0.33 mV peak-to-peak. What is the CMRR of the probe at 1 MHz?

4.11 Give specifications applicable to any type of probe.

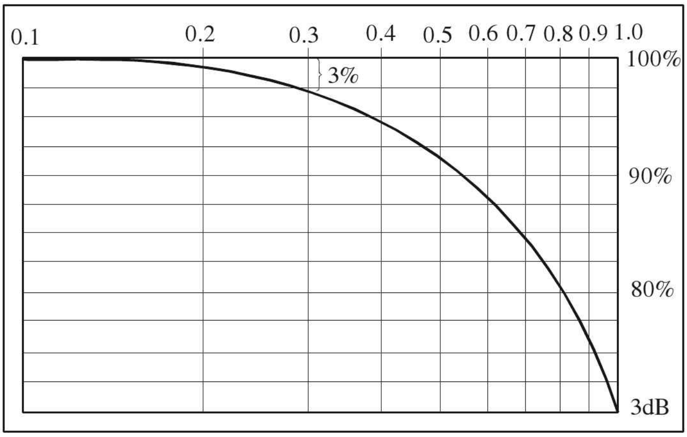


Figure Q4.8

Chapter 5

Bridge Instruments

5.1 INTRODUCTION

Bridge instruments are used to measure:

- resistance, inductance, and capacitance values of components
- purity of these components
- impedance and admittance of these components
- physical parameters like temperature, pressure, load, etc.

Various types of bridge circuits are employed in these instruments. These are categorized as dc and ac bridges. They operate on a null-indication principle. Therefore, the indication is independent of the calibration or any other characteristics of the indication device. It gives a very high degree of accuracy.

In this chapter we will study the basic bridge circuits, features of each circuit and their applications in the instruments used in measurement and control.

5.2 DIRECT CURRENT BRIDGES

The dc bridges are basically used for the measurement of resistance. Resistive elements are some of the most common sensors used in instrumentation. They can be made sensitive to temperature, strain (by pressure or by flex), and light. Using these basic elements, many complex physical phenomena can be measured, such as fluid or mass flow (by sensing the temperature difference between two calibrated resistances), dew-point humidity (by measuring two different temperature points), etc. The resistances of the sensor elements can range from less than 100 Ω to several hundred kΩ, depending on the sensor design and the physical environment to be measured. In some other applications, there may be a need to measure resistance values of as low as 10 Ω.

Wheatstone bridge and Kelvin bridge are the basic dc bridges. The Wheatstone bridge is used in two ways: (i) to measure the value of an unknown resistor by comparison with standard resistors, and (ii) to detect small changes in a resistance transducer.

5.2.1 Wheatstone Bridge

A basic Wheatstone bridge circuit consists of four resistors, a source of excitation *E*, and a voltage detector *G*, as illustrated in Figure 5.1. Two of the resistors, R_1 and R_2, are known and form the ratio arm, R_4 is unknown and R_3 is an adjustable standard resistor. The voltage detector measures the difference between the outputs of two voltage dividers connected across the excitation. The voltage detector may be a galvanometer, a CRO, any dc millivoltmeter or a digital display.

The current through the detector depends on the potential difference between the points c and d. The bridge is said to be balanced when this potential difference is zero and hence no current flows through the detector. Under this condition, we get the equation

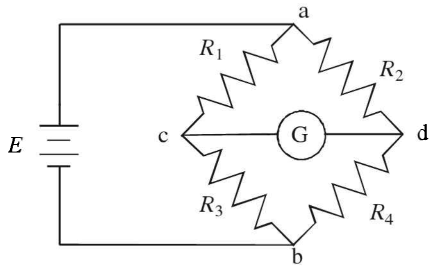


Figure 5.1 Wheatstone bridge.

$$\frac{R_1}{R_1+R_3}E = \frac{R_2}{R_2+R_4}E$$

from which we get

$$R_1R_4 = R_2R_3$$

Therefore an unknown resistance, R_4, can be expressed in terms of the other known resistances as

$$R_4 = R_3\frac{R_2}{R_1}$$

Resistor R_3 is the standard resistor, and R_2 and R_1 are the resistors of the ratio arms. R_3 is changed to balance the bridge. Note that the above equation is independent of the characteristics and calibration of the detector provided the detector

- observes the given unbalance,
- observes a minimum unbalance, and
- determines the expected deflection with the required degree of precision

We can define the *fractional uncertainty* in the measurement of unknown resistance R_4 by the equation

$$\frac{\Delta R_4}{R_4} = \sqrt{\left(\frac{\Delta R_1}{R_2}\right)^2 + \left(\frac{\Delta R_2}{R_2}\right)^2 + \left(\frac{\Delta R_3}{R_3}\right)^2}$$

where ΔR_i is the change in resistance R_i.

Example 5.1

The Wheatstone bridge has $R_1 = (1008 \pm 0.5)\ \Omega$ and $R_2 = (2002 \pm 0.5)\ \Omega$. At the balance point, $R_3 = (1080 \pm 5)\ \Omega$. Calculate the value of unknown resistance, R_4, and fractional uncertainty in its value.

Solution

At balance point,

$$R_4 = R_3 \frac{R_2}{R_1} = 1080 \times \frac{2002}{1008} = 2150\ \Omega$$

Fractional uncertainty in the measurement of R_4 is given by

$$\Delta R_4 = \sqrt{\left(\frac{\Delta R_1}{R_1}\right)^2 + \left(\frac{\Delta R_2}{R_2}\right)^2 + \left(\frac{\Delta R_3}{R_3}\right)^2} \times R_4$$

or

$$\Delta R_4 = \sqrt{\left(\frac{0.5}{1008}\right)^2 + \left(\frac{5}{1080}\right)^2 + \left(\frac{0.5}{2002}\right)^2} \times 2150$$

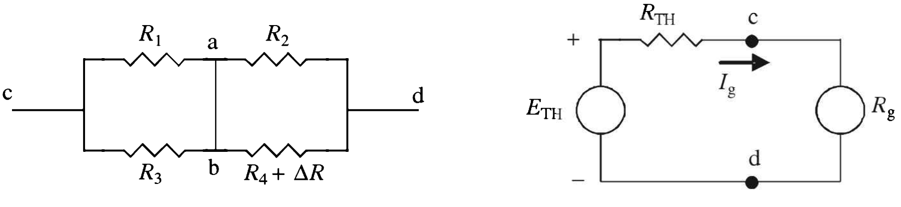
$$= 10\ \Omega$$

Therefore the unknown resistance can be represented as $(2150 \pm 10)\ \Omega$.

Sensitivity to unbalance—change in one resistance

Wheatstone bridge is also used to measure small changes in unknown resistor, R_4, for instance, due to temperature changes or the motion of microscopic defects in the resistor. As an example, suppose $R_4 = 10^6\ \Omega$ and we want to measure a change in R_4 of 1 Ω, resulting from a small temperature change. There is no ohmmeter which can reliably measure a change in resistance of 1 part in a million. However, the bridge can be set up so that the voltage between points c and d, $V_{cd} = 0$ when R_4 is exactly $10^6\ \Omega$. Then any change in R_4, i.e. ΔR, results in a non-zero V_{cd}, which, as we show below, is proportional to ΔR.

The sensitivity to unbalance of the bridge can be computed by analysing the bridge circuit for a small unbalance and then measuring the current through the detector. We will calculate the Thevenin voltage and Thevenin resistance of the bridge as seen by the detector in Figure 5.1. Figure 5.2(a) shows the Thevenin resistance, and Figure 5.2(b) gives the complete equivalent circuit with the detector.



(a) (b)

Figure 5.2 (a) Thevenin resistance circuit and (b) Thevenin equivalent circuit.

Let us assume the resistance of the battery to be zero and $\Delta R \ll R_4$. We thus get

$$E_{TH} = E\left(\frac{R_1}{R_1 + R_3} - \frac{R_2}{R_2 + R_4 + \Delta R}\right)$$

$$R_{TH} = \frac{R_1 R_3}{R_1 + R_3} + \frac{R_2(R_4 + \Delta R)}{R_2 + R_4 + \Delta R}$$

and

$$I_g = \frac{E_{TH}}{R_{TH} + R_g}$$

Case 1: Equal resistance bridge (i.e. $R_1 = R_2 = R_3 = R_4 = R$)
Then

$$E_{TH} = E\left(\frac{1}{2} - \frac{1}{2 + \frac{\Delta R}{R}}\right) \cong E\left(\frac{\Delta R}{4R}\right)$$

and

$$R_{TH} \cong R$$

Therefore,

$$I_g = \frac{E_{TH}}{R + R_g} = \frac{E\left(\frac{\Delta R}{4R}\right)}{R + R_g}$$

If we are using a galvanometer as a detector, then

Deflection of the galvanometer, $d = S_i . I_g = \dfrac{S_i\, E\left(\frac{\Delta R}{4R}\right)}{R + R_g}$

where S_i is the current sensitivity of the galvanometer measured in mm/μA.

We can define the sensitivity of the bridge as follows:

$$\text{Bridge sensitivity, } S_B = \frac{d}{\frac{\Delta R}{R}} = \frac{S_i E}{4(R + R_g)} = \frac{S_v E}{4}$$

where S_v is the voltage sensitivity of the galvanometer and is equal to $\dfrac{S_i}{R + R_g}$.

Case 2: Unequal resistance bridge

Here

$$E_{TH} = \frac{E\dfrac{\Delta R}{R_4}}{\left(1 + \dfrac{R_3}{R_1}\right)\left(1 + \dfrac{R_2}{R_4} + \dfrac{\Delta R}{R_4}\right)}$$

$$= \frac{E\dfrac{\Delta R}{R_4}}{\left(1 + \dfrac{1}{n}\right)\left(1 + n + \dfrac{\Delta R}{R_4}\right)}$$

where $n = R_2/R_4 = R_1/R_3$.

Let $p = \dfrac{\Delta R}{R_4}$ and $p << n + 1$. We therefore have

$$E_{TH} = \frac{E\,p\,n}{(1 + n)^2} \tag{5.1}$$

Figure 5.3 shows the plot of E_{TH}/E versus p, i.e. change in R_4 (i.e. ΔR) for different values of n.

From Figure 5.3 we observe that the bridge output and hence the sensitivity of the bridge is maximum for $n = 1$ and reduces with increase in n, where $n = R_2/R_4 = R_1/R_3$.

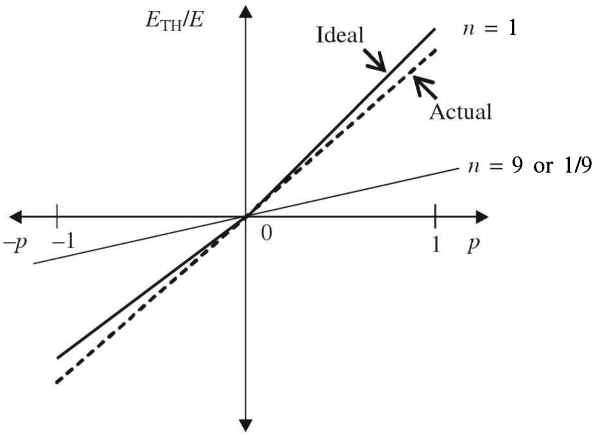


Figure 5.3 Plot of E_{TH}/E versus p.

Further, we infer that since in the expression E_{TH}/E, p is multiplied by $n/(1 + n)^2$, the linearity between E_{TH}/E and p improves as n increases. It is the poorest for $n = 1$. So, we compromise the value of n for better linearity and sensitivity.

We can express the deflection of the galvanometer, d, as follows:

$$d = S_i . I_g = \frac{S_i . E\, p\, n}{(R + R_g)(1 + n)^2}$$

and

$$\text{Bridge sensitivity, } S_B = \frac{d}{p} = \frac{S_v . E . n}{(1 + n)^2}$$

Maximum sensitivity is for $n = 1$ and is equal to $S_v . E/4$. As the value of n increases or decreases, S_B decreases. This reduces the accuracy of balancing of the bridge.

Example 5.2

The Wheatstone bridge, as shown in Figure 5.1, having $R_1 = 1000\ \Omega$ and $R_2 = 100\ \Omega$, is used to measure an unknown resistance of 25 Ω. Two galvanometers are available. Galvanometer A has a sensitivity of 200 mm/μA and internal resistance of 20 Ω and Galvanometer B has a sensitivity of 600 mm/μA and internal resistance of 600 Ω. Which of the galvanometers is more sensitive to a small unbalance in the bridge?

Solution

Let us find the standard resistance, R_3, at balance point.

$$R_3 = R_4 \frac{R_1}{R_2} = 25 \times \frac{1000}{100} = 250\ \Omega$$

Therefore, Thevenin resistance is given by

$$R_{TH} = \frac{R_1 R_3}{R_1 + R_3} + \frac{R_2 R_4}{R_2 + R_4} = \frac{1000 \times 250}{1000 + 250} + \frac{100 \times 25}{100 + 25} = 220\ \Omega$$

The deflection of the galvanometer, d, is given by

$$d = S_i . I_g = \frac{S_i . E\, p\, n}{(R_{TH} + R_g)(1 + n)^2}$$

where S_i is the sensitivity of the galvanometer, $p = \Delta R/R_4$, $n = R_2/R_4$ and R_g is the resistance of the galvanometer.

If d_A is the deflection of galvanometer A and d_B is the deflection of galvanometer B, then

$$\frac{d_A}{d_B} = \frac{S_{iA}}{S_{iB}} \times \frac{R_{TH} + R_{gB}}{R_{TH} + R_{gA}} = \frac{200}{600} \times \frac{220 + 600}{220 + 20} = 2.28$$

Therefore, galvanometer A is 2.28 times more sensitivity than galvanometer B.

Example 5.3

A Wheatstone bridge can be used to detect location of ground faults in cables. This test is called *Varley Loop* test. The setup is shown in Figure 5.4 where a cable is connected between b and d. Let $R_A + R_B$ be the resistance of the cable. Let the ground fault in the cable occur at a distance X from the bridge. The bridge is first balanced with switch S at position 1 with $R_3 = 100\ \Omega$. Then the switch is changed to position 2 and balance is achieved with $R_3 = 99\ \Omega$. If the resistance of the cable is 0.15 Ω/km, find the distance X of the ground fault from the bridge.

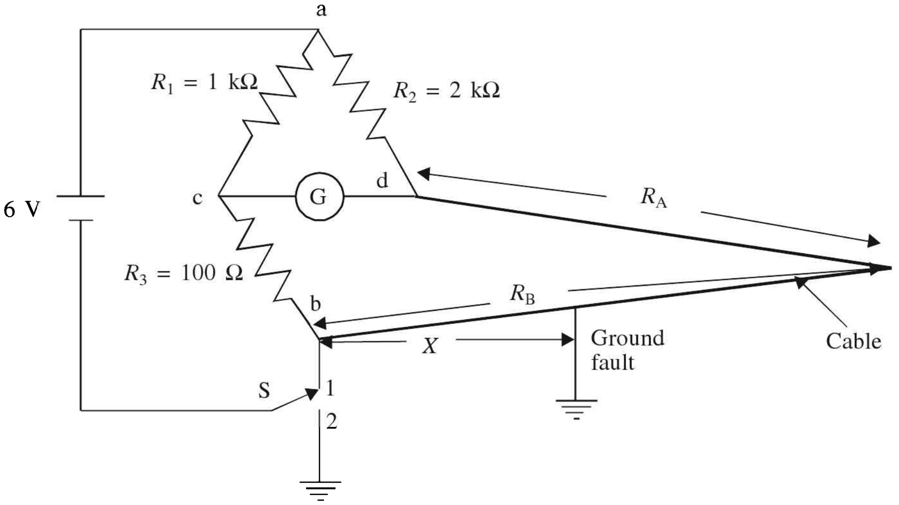


Figure 5.4 Varley loop test.

Solution

With switch at position 1, the balance equation is

$$R_A + R_B = R_3 \frac{R_2}{R_1} = 100 \times \frac{2}{1} = 200\ \Omega$$

With switch at position 2, the balance equation is

$$R_A + R_B - R_X = (R_3 + R_X)\frac{R_2}{R_1} = (99 + R_X) \times \frac{2}{1}$$

Therefore,

$$R_X = \frac{2}{3}\ \Omega$$

and distance X of the ground fault from the bridge is equal to

$$X = \frac{2/3}{0.15}\ \text{km} = 4444.4\ \text{m}$$

Example 5.4

The Wheatstone bridge of Figure 5.1 has battery voltage = 20 V and $R_1 = R_2 = R_3 = 1000\ \Omega$. R_4 is a thermistor, which is a resistor whose value is very sensitive to temperature and has negative temperature coefficient. The temperature coefficient of the thermistor is 5 Ω/°C. At 50°C the resistance of the thermistor is 1000 Ω. At what temperature is the Thevenin voltage of the bridge equal to 50 mV?

Solution

For equal resistance bridge, Thevenin voltage is given by

$$E_{\text{TH}} \cong E\left(\frac{\Delta R}{4R}\right)$$

i.e.

$$E_{\text{TH}} = 0.05 = 20\left(\frac{\Delta R}{4000}\right)$$

Therefore,

$$\Delta R = 0.05 \times 200 = 10\ \Omega$$

For 10 Ω increase in resistance, the temperature of the thermistor decreases by 2°C. Hence the new temperature is 48°C.

Improving linearity and sensitivity

From Figure 5.3, we see that to improve upon the sensitivity, we should use the ratio of R_2 and R_4, and R_1 and R_3, i.e. n to be equal to one. Resistive transducers designed in pairs are used in critical applications in instrumentation. They also provide temperature compensation if we choose the resistors in push–pull pairs ($n = 1$), i.e. the resistance in one arm increases whereas it decreases in the other. If p_1 and p_2 correspond to fractional changes in R_4 and R_2 respectively, then

$$E_{\text{TH}} = \frac{E(p_1 + p_2)}{2(2 + p_1 - p_2)}$$

For $p_1 = p_2 = p$, $E_{\text{TH}} = E.p/2$.

This is double the voltage without using push–pull pairs. So sensitivity is doubled. Further, E_{TH} varies linearly with p.

The sensitivity can be increased further by increasing E, but this may heat up the resistors thereby changing the resistance value.

Example 5.5

The Wheatstone bridge of Figure 5.1, has battery voltage = 6 V and $R_1 = R_3 = 120\ \Omega$. R_2 and R_4 are two strain gauges each of 120 Ω resistance. The gauge factor of the strain gauge is 2.0. The strain of value 0.000165 is applied to the gauges in such a way that one is in tension and the other in compression. If the battery current in the initial balance condition is 100 mA, determine (a) the output voltage of the bridge and (b) the sensitivity in volt per unit strain.

Solution

The output voltage of the bridge in balance condition is given by

$$E_{\mathrm{TH}} = E\left[\frac{R+\Delta R}{(R+\Delta R)+(R-\Delta R)} - \frac{1}{2}\right] = \frac{\Delta R}{2R}E$$

$$= \frac{G\varepsilon}{2}E$$

where G is the strain gauge factor and ε is the strain.
Therefore,

$$E_{\mathrm{TH}} = \frac{2 \times 0.000165 \times 6}{2} = 0.99 \text{ mV}$$

and sensitivity is given by

$$\frac{E_{\mathrm{TH}}}{\varepsilon} = \frac{0.99 \times 10^{-3}}{0.000165} = 6.0 \text{ V/strain}$$

Excitation of the bridge with constant current source

We can excite the bridge with a constant current source instead of a constant voltage source as explained above. All the above relations will change.

Figure 5.5 shows the bridge with a constant current source excitation. We can analyse the circuit first by finding the voltage between points a and b, E_{ab} under bridge balance condition and then calculating the Thevenin voltage, E_{TH}.

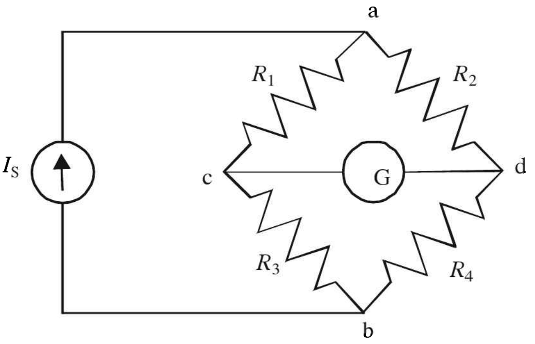


Figure 5.5 Wheatstone bridge with a constant current source excitation.

$$E_{\mathrm{ab}} = I_S\left[(R_1 + R_3)\,||\,(R_2 + R_4)\right]$$

$$= I_S\left[\frac{mR_4(n+1)}{1+m}\right]$$

where $n = R_2/R_4 = R_1/R_3$ and $m = R_3/R_4 = R_2/R_1$.

From Eq. (5.1) substituting E with E_{ab}, we get

$$E_{TH} = I_S\left(\frac{m}{m+1}\right)\left(\frac{n}{n+1}\right)pR_4 = I_S\left(\frac{m}{m+1}\right)\left(\frac{n}{n+1}\right)\Delta R \tag{5.2}$$

Note that E_{TH} is proportional to pR_4, i.e. ΔR rather than p in constant voltage excitation. Sensitivity $E_{TH}/\Delta R$ increases with n and m. Figure 5.6 shows the variation of $E_{TH}(m+1)/(mI_S)$ with ΔR for different values of n.

The constant current bridge is normally preferred with low resistance values that need to be operated at constant current.

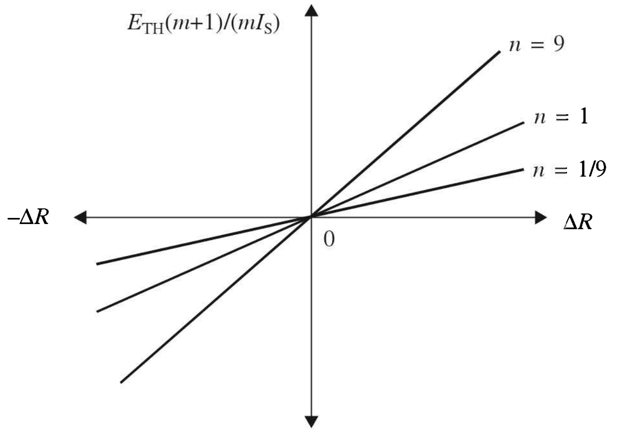


Figure 5.6 Variation of $E_{TH}(m+1)/(mI_S)$ with ΔR for different values of n.

Sensitivity to unbalance—change in all resistances

From Figure 5.1, we can write the voltage across the detector under balance conditions as

$$E_g = E_{cd} = \frac{R_1}{R_1+R_3}E - \frac{R_2}{R_2+R_4}E$$
$$= \left[\frac{R_1R_4 - R_2R_3}{(R_1+R_3)(R_2+R_4)}\right]E$$

Now suppose that all resistances change during the measurement. The corresponding voltage reading will be

$$E_g + \Delta E_g = \frac{(R_1+\Delta R_1)(R_4+\Delta R_4) - (R_2+\Delta R_2)(R_3+\Delta R_3)}{(R_1+\Delta R_1+R_3+\Delta R_3)(R_2+\Delta R_2+R_4+\Delta R_4)}E$$

If the bridge is initially balanced, the initial voltage reading E_g should be zero and

$$n = \frac{R_2}{R_4} = \frac{R_1}{R_3}$$

Using this, we get the change in E_g as

$$\Delta E_g = \frac{n}{(1+n)^2}\left[\frac{\Delta R_1}{R_1} - \frac{\Delta R_3}{R_3} + \frac{\Delta R_2}{R_2} - \frac{\Delta R_4}{R_4}\right](1+\eta)E$$

where η is given by

$$\eta = \cfrac{1}{1 + \cfrac{1+n}{\frac{\Delta R_1}{R_1} + \frac{\Delta R_4}{R_4} + n\left[\frac{\Delta R_2}{R_2} + \frac{\Delta R_3}{R_3}\right]}}$$

For small changes in the resistances, $\eta = 0$. Therefore,

$$\Delta E_g = \frac{n}{(1+n)^2}\left[\frac{\Delta R_1}{R_1} - \frac{\Delta R_3}{R_3} + \frac{\Delta R_2}{R_2} - \frac{\Delta R_4}{R_4}\right]E$$

Further, if we choose the same resistance value for all four resistors, $R_1 = R_2 = R_3 = R_4 = R$, and noting that $n = 1$ in this case, the change in voltage can be simplified to

$$\Delta E_g = \frac{\Delta R_1 - \Delta R_3 + \Delta R_2 - \Delta R_4}{R}E$$

By properly selecting the resistances, the Wheatstone bridge circuit can amplify small changes in resistance and/or compensate for changes in temperature.

Calibration of the bridge

Figure 5.7 shows the arrangement of a bridge which has the provision for the sensitivity, zero adjustment and calibration for full-scale reading.

Normally E is a stabilised power source. To compensate the voltage drop across the lead wires, sensitivity variable resistance is provided. Zero adjustment variation is used for adjusting zero when the bridge is balanced. For full-scale reading on the detector, a variable resistance for calibration is provided. Initially, switch S is open and the bridge is balanced. When the switch is closed, a known resistance R_{sh} comes in parallel with R_2. The desired value of the detector voltage is then observed.

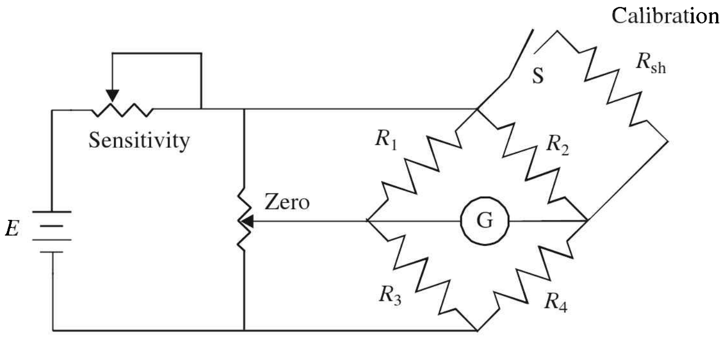


Figure 5.7 Bridge with various adjustments.

Limitations of the Wheatstone bridge

Wheatstone bridge can be used to measure resistance values ranging from 1 Ω to 1 MΩ with an accuracy of approximately 0.1%. However, the following difficulties are encountered when very high and very low resistances are measured.

- Resistances less than 1 Ω are difficult to measure accurately because of uncertainty arising from the contact resistance present between the resistor to be measured and the binding posts of the bridge, and also the resistance of the connecting leads. The resistance of the leads can be calculated or measured and the final result modified, but contact resistance is difficult to measure or compute. One of the most elementary precautions concerning the use of a bridge when measuring low resistance, is to tighten the binding posts securely so that the contact resistance between the binding posts and the resistance to be measured is minimum.
- Thermoelectric effect can also cause problems especially for low value resistance measurements. To reduce this resistance, contact of dissimilar metals with one another should be avoided so that no thermal emf is generated.
- For low value of the resistance, the change in the resistance value with temperature may be an appreciable part of the resistance value and can give a wrong reading. Therefore, the temperature coefficient of the resistance to be measured should be very low.
- Measurement of resistances greater than 1 MΩ becomes difficult because the voltage applied to the bridge must be substantially increased to obtain definite detection of the output. The result is that an increase in the supply voltage increases the power dissipation (heat) of the bridge resistors. The change in resistance, because of the heat, is sufficient to produce an appreciable error. Further, when the resistance values are so high that they approach insulation resistance, leakage in the insulation will increase inaccuracy. Leakage paths between the resistor leads along the outside surface of the resistor body must be avoided when resistances greater than 1 MΩ are measured.

Guarded Wheatstone bridge

One of the major problems of an ordinary dc Wheatstone bridge in measurements of high resistance is the leakage that takes place over or around the component being measured, or over the binding posts by which component is attached to the instrument, or within the instrument itself. These currents also vary with the humidity level of the environment. Some guard circuits are used to reduce or remove the leakage paths.

Using guard wire. Figure 5.8 shows a simple guard circuit in the unknown resistance arm of the bridge. A guard wire completely surrounds the surface of the insulated post. In the absence of the guard wire, the leakage current I_L along the insulated surface of the binding adds to the current I_X thereby giving a wrong reading. Guard wire post intercepts the leakage current and returns it to the battery. Note that the guard does not touch any part of the bridge circuitry and is directly connected to the battery. This principle can be applied for any part of the bridge where leakage affects the measurement.

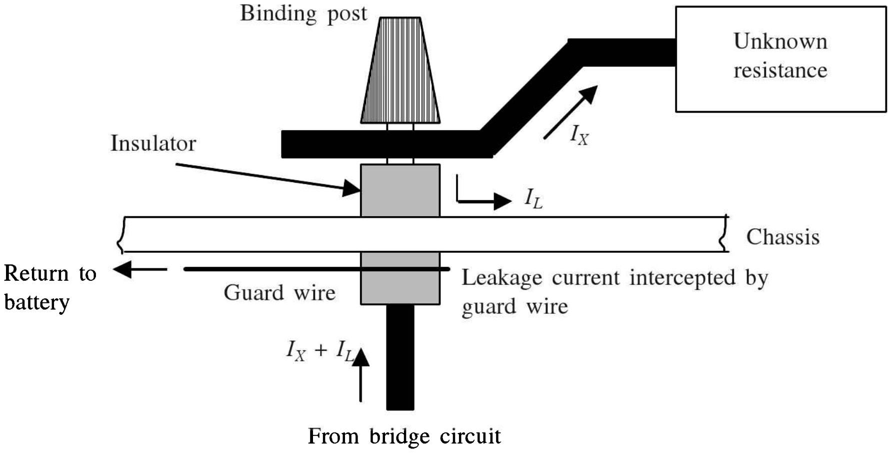


Figure 5.8 Use of guard wire to eliminate leakage current.

Using three terminal resistance. This technique is used to avoid the effect of leakage current external to the bridge circuitry. The junction of R_1 and R_3 is brought out as a separate guard terminal on the front panel of the instrument and is used to connect a three terminal resistance formed by the binding posts where the unknown resistor, R_X is connected as shown in Figure 5.9.

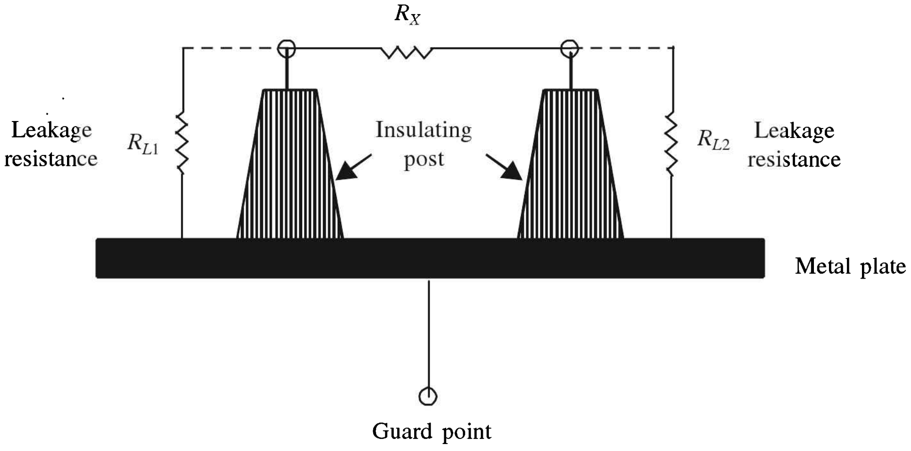


Figure 5.9 Three terminal resistance.

Figure 5.10 shows the connection of a three terminal resistance to the bridge. Resistor R_{L1} is in parallel with R_3 and R_{L2} is in parallel with detector, G. Since R_{L1} and R_{L2} are very large compared to R_3 and the resistance of the detector, their effect is negligible. The effect of the external leakage paths is therefore removed by this connection.

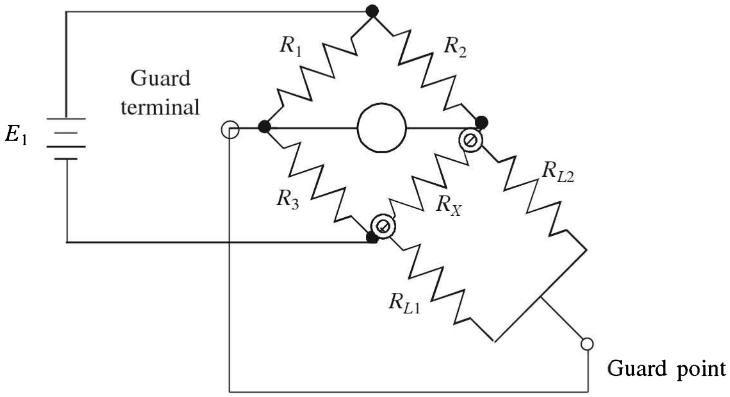


Figure 5.10 Three terminal resistance connected to the bridge.

Remote connected Wheatstone bridge

Wiring resistance and noise pickup are the biggest problems associated with remotely located bridges. For example, consider a 350 Ω strain gauge (R_4), as shown in Figure 5.11, connected to the bridge circuit, with values of other arms resistors as $R_1 = R_2 = R_3 = 350$ Ω, by 100 feet of 30 gauge twisted pair copper wire. The resistance of the wire, R_{LEAD}, at 25°C is 0.105 Ω/ft, or 10.5 Ω for 100 ft. The total lead resistance in series with the 350 Ω strain gauge is therefore 21 Ω. To compensate for lead resistance a compensation resistance, R_{COMP} has been added in series with R_3 to balance the bridge. Without R_{COMP}, the bridge has an output offset voltage of 145.63 mV for a nominal strain gauge resistance of 350 Ω.

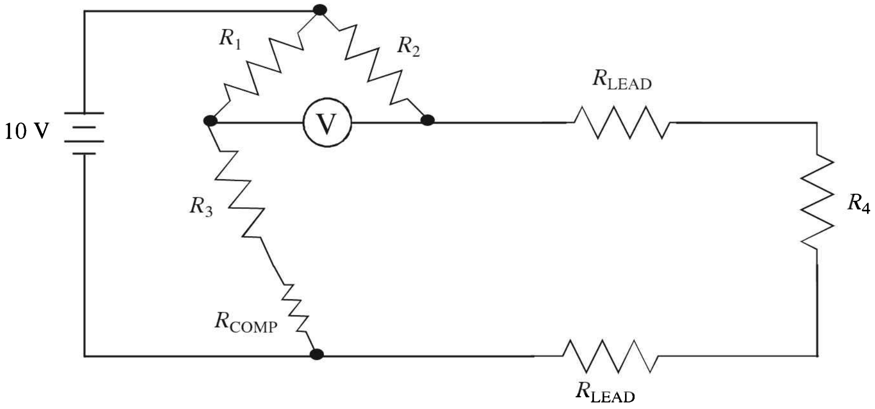


Figure 5.11 Remotely connected bridge.

Let the full-scale variation of the strain gauge resistance above its nominal 350 Ω value be +1% (+3.5 Ω). This corresponds to a full-scale strain gauge resistance of 353.5 Ω which causes a bridge output voltage of +23.45 mV.

Let the temperature coefficient of the copper wire be 0.385%/°C. Assume that the cable temperature increases by +10°C above the nominal room temperature. This results in a total lead resistance increase of +0.404 Ω (10.5 Ω × 0.00385/°C × 10°C) in each lead. The total additional lead resistance (of the two leads) is +0.808 Ω. With no strain, this additional lead resistance produces an offset of +5.44 mV in the bridge output. Full-scale strain produces a bridge output of +28.83 mV (a change of +23.39 mV from no strain). Thus the increase in temperature produces an offset voltage error of +5.44 mV (+23% full-scale) and a gain error of −0.06 mV (23.39 mV − 23.45 mV), or −0.26% full-scale. Note that these errors are produced solely by the 30 gauge wire, and do not include any temperature coefficient errors in the strain gauge itself.

The effects of wiring resistance on the bridge output can be minimized by the three-wire connection as shown in Figure 5.12. We assume that the bridge output voltage is measured by a high impedance device, therefore there is no current in the sense lead. Note that the sense lead measures the voltage output of a divider: the top-half is the bridge resistor plus the lead resistance, and the bottom-half is the strain gauge resistance plus the lead resistance. The nominal sense voltage is therefore independent of the lead resistance. When the strain gauge resistance increases to full-scale (353.5 Ω), the bridge output increases to +24.15 mV.

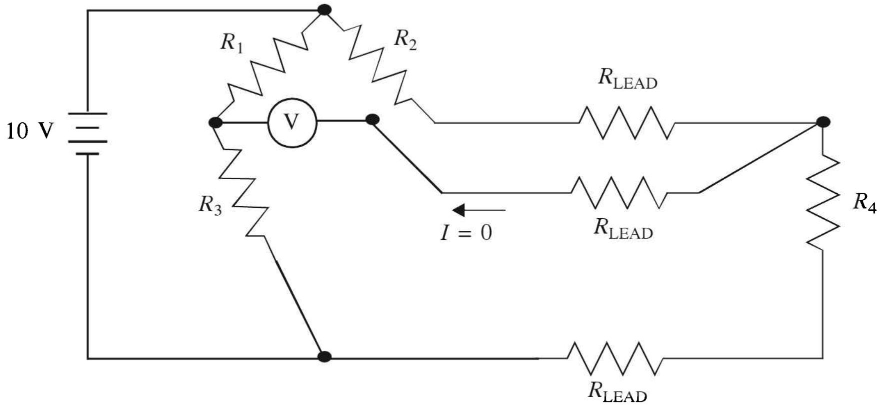


Figure 5.12 Three-wire connection to a remotely connected bridge.

Increasing the temperature to +35°C increases the lead resistance by +0.404 Ω in each half of the divider. The full-scale bridge output voltage decreases to +24.13 mV because of the small loss in sensitivity, but there is no offset error. The gain error due to the temperature increase of +10°C is therefore only −0.02 mV, or −0.08% of full-scale which is much smaller than the +23% full-scale offset error and the −0.26% gain error as for the two-wire connection shown in Figure 5.11.

The three-wire method works well for remotely located resistive elements which make up one leg of a single-element varying bridge. However, all-element varying bridges, in which all arm resistances vary, generally, are housed in a complete assembly as in the case of a load cell. When these bridges are remotely located from the conditioning electronics, special techniques are used to maintain accuracy. Of particular concern is maintaining the accuracy and stability of the bridge excitation voltage. Since the bridge output is directly proportional to the excitation voltage, any drift in the excitation voltage produces a corresponding drift in the output voltage.

For this reason, most all-element varying bridges (such as load cells) are six-lead assemblies: two leads for the bridge output, two leads for the bridge excitation, and two sense leads. This method (called four-wire sensing) is shown in Figure 5.13.

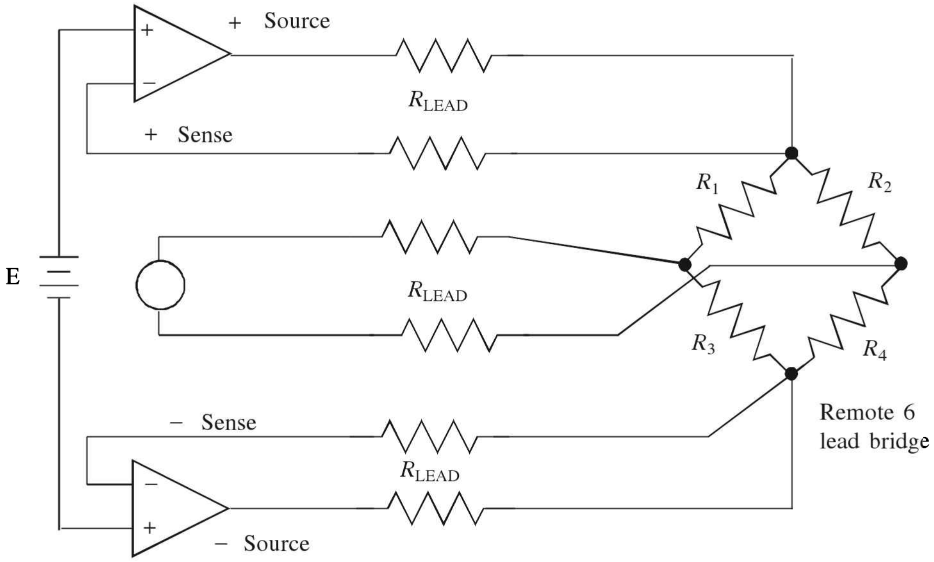


Figure 5.13 Four-wire sensing.

The sense lines go to high impedance operational amplifier (opamp) inputs, thus there is minimal error due to the bias current induced voltage drop across their lead resistance. The opamps maintain the required excitation voltage to make the voltage measured between the sense leads always equal to E. Although four-wire sensing eliminates errors due to voltage drops in the wiring resistance, the drive voltages must still be highly stable since they directly affect the bridge output voltage. In addition, the opamps must have low offset, low drift, and low noise.

A very powerful *ratiometric* technique which includes four-wire sensing to minimize errors due to wiring resistance and also eliminates the need for an accurate excitation voltage, can be designed using analog-to-digital converter (ADC) (e.g. AD 7730) having differential reference input and differential analog input terminals. ADC is driven from a single supply voltage which is also used to excite the remote bridge. Both the analog input and the reference input to the ADC are high impedance and fully differential. By using the + and − SENSE outputs from the bridge as the differential reference to the ADC, there is no loss in measurement accuracy if the actual bridge excitation voltage varies. The AD 7730 is one of a family of sigma-delta ADCs with high resolution (24 bits) and internal programmable gain amplifiers (PGAs) and is ideally suited for bridge applications. These ADCs have self- and system-calibration features which allow offset and gain errors due to the ADC to be minimized. For instance, the AD 7730 has an offset drift of 5 nV/°C and a gain drift of 2 ppm/°C. Offset and gain errors can be reduced to a few microvolts using the system calibration feature.

The constant current excitation method is another method for minimizing the effects of wiring resistance on the measurement accuracy. However, the accuracy of the reference, the sense resistor, and the opamp—all influence the overall accuracy.

5.2.2 Kelvin Bridge

A Kelvin bridge is used for measuring resistances from microohms to 1 ohm. In Wheatstone bridge this is not possible due to lead wire resistance.

Let us see the effect of connecting leads on the measurements. Consider the Wheatstone bridge circuit as shown in Figure 5.14. Resistance R_y (also called yoke resistance) represents the lead resistance from R_2 to R_x. When detector G is connected to point m, the resistance R_y is added to unknown resistance, increasing the value of R_x. When G is connected to point n, the resistance R_y is added to R_2 thereby decreasing the value of R_x. Let G be connected to the point p between m and n such that

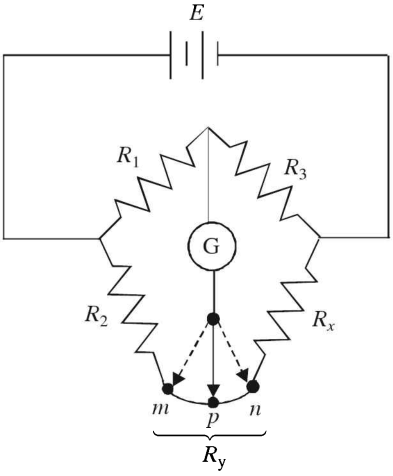


Figure 5.14 Wheatstone bridge showing lead resistance R_y.

$$\frac{R_{np}}{R_{mp}} = \frac{R_3}{R_1}$$

The bridge balance equation then becomes

$$R_x + R_{np} = \frac{R_3}{R_1}(R_2 + R_{mp})$$

From these two equations, we get

$$R_x = \frac{R_3}{R_1}R_2$$

which is the usual balance equation of the Wheatstone bridge. We see that the effect of lead resistance from point *m* to *n* has been eliminated by connecting the detector to intermediate position *p*. But practically it may be difficult to identify the point *p* correctly. Instead, we can use two resistors with correct ratio between point *m* and the detector and point *n* and the detector. This forms Kelvin double bridge (also called Kelvin bridge).

Kelvin double bridge

Figure 5.15 shows the circuit diagram of a Kelvin double bridge. Since the circuit contains two ratio arms (R_1 and R_3 and a and b), it is called the double bridge. The detector reading is zero when the voltages between *kl* and *lmp* are equal. That is

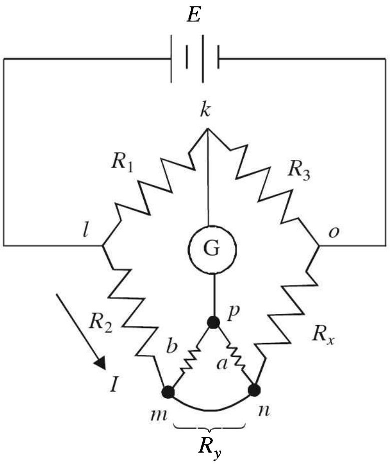


Figure 5.15 Kelvin double bridge.

$$E_{kl} = E_{lmp}$$

Now

$$E_{kl} = \frac{R_1}{R_1 + R_3} E$$

$$= \frac{R_1}{R_1 + R_3} I \left[R_2 + R_x + \frac{(a+b)R_y}{a+b+R_y} \right]$$

and

$$E_{lmp} = IR_2 + E_{mn} \frac{b}{a+b}$$

where

$$E_{mn} = I \frac{(a+b)R_y}{a+b+R_y}$$

Therefore

$$E_{lmp} = I\left[R_2 + \frac{b}{a+b}\frac{(a+b)R_y}{a+b+R_y}\right]$$

For $E_{kl} = E_{lmp}$, we have

$$R_x = \frac{R_3}{R_1}R_2 + \frac{bR_y}{a+b+R_y}\left(\frac{R_3}{R_1} - \frac{a}{b}\right)$$

If $\frac{a}{b} = \frac{R_3}{R_1}$, then

$$R_x = \frac{R_3}{R_1}R_2$$

which is the usual balance equation of the Kelvin bridge and is independent of R_y, the lead resistance.

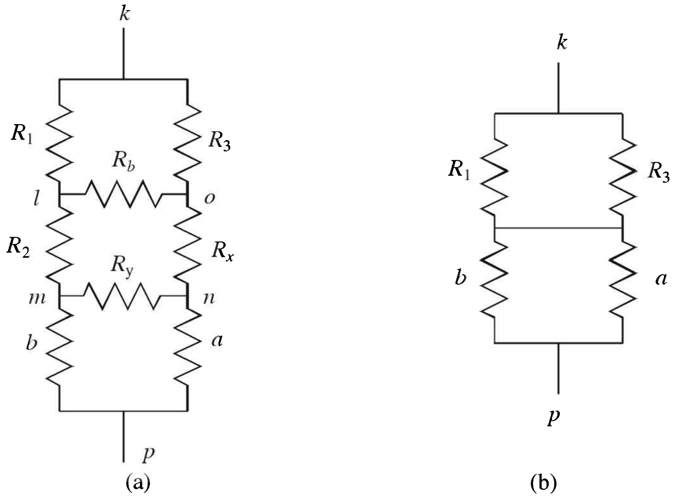
So, the lead resistance (or yoke resistance) has no effect on the measurement, provided that the two sets of ratio arms have equal resistance ratios.

Unbalanced Kelvin bridge

The detector current under unbalanced conditions can be found by finding the Thevenin voltage and Thevenin resistance across it.

Thevenin voltage across G, $E_{TH} = E_{kl} - E_{lmp}$

To find Thevenin resistance, let us draw the equivalent circuit across G, as shown in Figure 5.16(a). R_b is the battery resistance. Since R_2, R_x, R_b and R_y are very small, these can be neglected. The equivalent circuit then is shown in Figure 5.16(b).



(a) (b)

Figure 5.16 Thevenin resistance equivalent circuit.

Thevenin resistance is therefore equal to

$$R_{\mathrm{TH}} = \frac{R_1R_3}{R_1+R_3} + \frac{ab}{a+b}$$

and the detector current, I_g, is given by

$$I_g = \frac{E_{\mathrm{TH}}}{R_{\mathrm{TH}} + R_g}$$

where R_g is the detector resistance.

Example 5.6

A Kelvin double bridge is balanced with $b = 1200\ \Omega = 1600a$, $R_1 = 800a = 1.25\ R_2$. What is the value of R_x?

Solution

Under balanced condition,

$$R_x = \frac{R_3}{R_1}R_2 = \frac{a}{b}R_2$$

We are given

$$\frac{a}{b} = \frac{1}{1600},\ a = \frac{1200}{1600},\ R_2 = \frac{800a}{1.25} = \frac{600}{1.25} = 480$$

Therefore

$$R_x = \frac{480}{1600} = 0.3\ \Omega$$

Example 5.7

In a Kelvin double bridge, $R_1 = R_3 = a = b = 1000\ \Omega$. The battery has a value of 100 volts and a resistance of 2 Ω is connected in series with the battery. The resistance of the galvanometer is 250 Ω. Neglect the resistance of the link connecting the unknown resistance R_x to the standard resistance R_2. The bridge is balanced when $R_2 = 0.001\ \Omega$. Assuming galvanometer sensitivity as 300 mm/μA, calculate the deflection on the detector when the unknown resistance changes by 0.1% from its value at balance.

Solution

Refer to Figure 5.14. Under the balanced condition, the value of R_x is equal to

$$R_x = \frac{R_3}{R_1}R_2 = \frac{1000}{1000}0.001 = 0.001\ \Omega$$

With 0.1% change in R_x, the new value of R_x is

$$R_{x,\mathrm{new}} = 0.001 \times 1.001 = 0.001001\ \Omega$$

Since a resistance of 2 Ω is connected in series with the battery, the voltage across the points *l* and *o* is not equal to 100 V. It is given by

$$E_{lo} = \frac{R_2 + R_x + R_y}{R_2 + R_x + R_y + 2} \times 100 = 100 \text{ mV}$$

Thevenin voltage across G,

$$E_{\text{TH}} = E_{kl} - E_{lmp} = \left[\frac{R_3}{R_1 + R_3} - \frac{R_{x,\text{new}}}{R_{x,\text{new}} + R_2} \right] \times E_{lo}$$

$$= 0.025 \text{ mV}$$

Thevenin resistance across G is equal to

$$R_{\text{TH}} = \frac{R_1 R_3}{R_1 + R_3} + \frac{ab}{a+b} \times \frac{1000 \times 1000}{1000 + 1000} + \frac{1000 \times 1000}{1000 + 1000} = 1000 \ \Omega$$

The galvanometer current, I_g, is given by

$$I_g = \frac{E_{\text{TH}}}{R_{\text{TH}} + R_g} = \frac{0.025 \times 10^{-3}}{1000 + 250} = 0.02 \ \mu\text{A}$$

Therefore the galvanometer deflection, *d*, is equal to

$$d = 300 \times 0.02 = 6 \text{ mm}$$

A practical circuit of Kelvin double bridge

A more practical circuit of Kelvin double bridge is shown in Figure 5.17. The circuit reduces the contact potential drops in the measuring circuit. The standard resistor has been split into

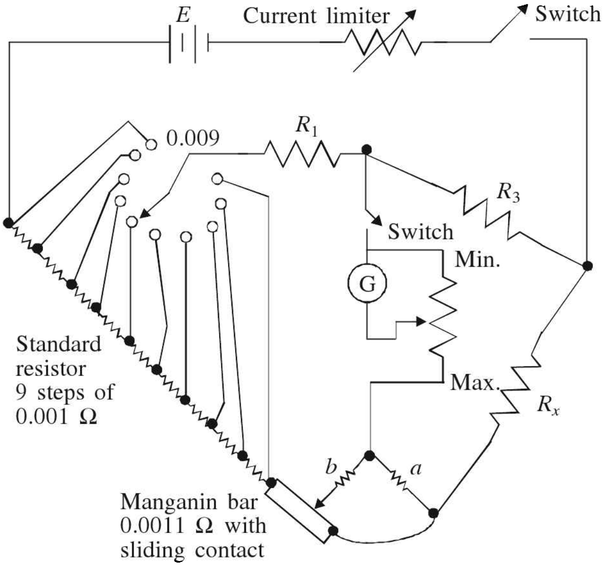

E
Current limiter
Switch
0.009
R_1
R_3
Switch
Min.
G
Max.
R_x
Standard resistor 9 steps of 0.001 Ω
b
a
Manganin bar 0.0011 Ω with sliding contact

Figure 5.17 A practical circuit of Kelvin double bridge.

nine steps of 0.001 Ω and a series resistance of calibrated manganin bar of 0.0011 Ω with a sliding contact for balancing. The total resistance of R_2 is therefore 0.0101 Ω. When the switch position is changed to select a suitable value of the standard resistor, the resistor of R_1 arm changes thereby changing the ratio of R_3 and R_1. But the total resistance seen by the battery does not change. The contact resistance comes in series with a high resistance of ratio arms, and so has negligible effect. The accuracy of measurement of R_x increases by selecting the ratio R_3/R_1 so that a large part of the standard resistor is used in the measuring circuit.

This circuit is used for the measurement of winding resistances of machines and transformers, contact and earth conductor resistances with an accuracy of about ±0.2%.

5.3 ALTERNATING CURRENT BRIDGES

The ac bridges are usually employed to measure impedance, inductance, capacitance, quality factor and dissipation factor conveniently and accurately. They are also used for phase shifting, providing feedback paths for oscillators and amplifiers, filtering out undesirable signals and measuring frequency.

Like a dc bridge, an ac bridge in its basic form consists of four arms (each an impedance), a source of excitation (ac source) and a balance detector sensitive to small alternating potential differences like head phone, vibration galvanometer or tuneable amplifier detector.

Before we discuss the bridge, let us familiarize ourselves with the impedance and other related terms used in these bridges.

Impedance is the basic electrical parameter used to characterize electronic circuits, components, and materials. It is defined as the ratio of the voltage applied to the device and the resulting current through it. That is, impedance is the total opposition a circuit offers to the flow of an alternating current (ac) at a given frequency, and is generally represented as a complex quantity, Z. Resistance (R), inductance (L) and capacitance (C) form an impedance.

$$\text{Impedance, } Z = \frac{E}{I} = R + jX$$

where X is defined as the reactance.

In real world, electronic components are not pure resistors, inductors or capacitors, but a combination of all three or two as shown in Figure 5.18. The figure also shows the phasor diagrams of each of the combinations. Today's generation of LCR bridges are capable of displaying these parameters and can easily calculate and display many other parameters such as Z, Y, X, G, B, D, etc. which are defined in Table 5.1.

Here f = frequency in hertz, $\omega = 2\pi f$, R and X are equivalent series quantities unless otherwise defined, G and B are equivalent parallel quantities unless otherwise defined. Subscript 'p' stands for parallel and subscript 's' stands for series. We define Q as being positive if it is inductive, negative if it is capacitive. We define D as positive if capacitive. The abbreviation ESR stands for Equivalent Series Resistance. A low value of D or a high value of Q means capacitor or inductor is quite pure while a high D or low Q means resistor is nearly pure. Note that D or Q of an impedance is independent of the configuration of the equivalent circuit used to represent it.

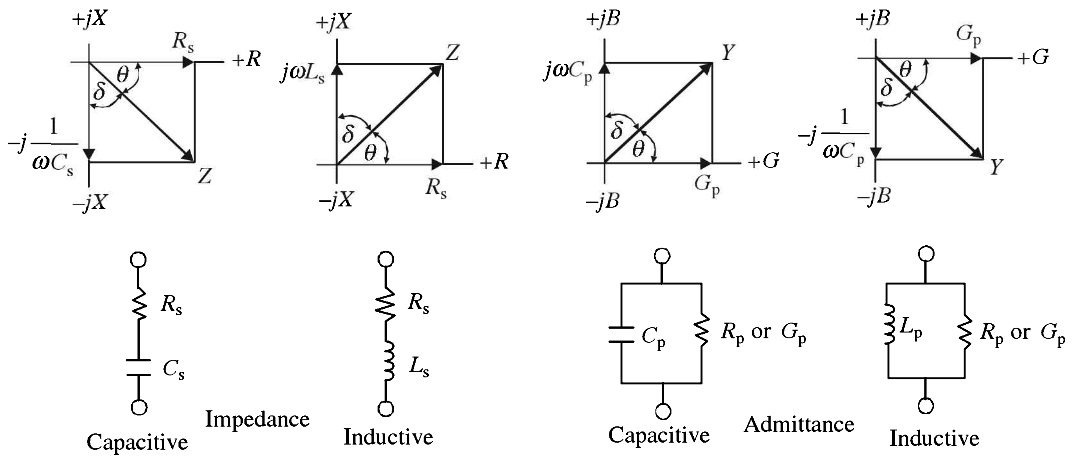


Figure 5.18 Phasor diagrams of series and parallel representation of components.

Table 5.1 Impedance and associated terms

Parameter	*Quantity*	*Unit, symbol*	*Formula*
Z	Impedance	ohm, Ω	$Z = R_s + jX_s = \frac{1}{Y} = \lvert Z \rvert e^{j\theta}$
$\lvert Z \rvert$	Magnitude of Z	ohm, Ω	$\lvert Z \rvert = \sqrt{R_s^2 + X_s^2} = \frac{1}{\lvert Y \rvert}$
R_s or ESR	Resistance, real part of Z	ohm, Ω	$R_s = \frac{G_p}{G_p^2 + B_p^2}$
X_s	Reactance, imaginary part of Z	ohm, Ω	$X_s = \frac{B_p}{G_p^2 + B_p^2}$
Y	Admittance	siemens, S	$Y = G_p + jB_p = \frac{1}{Z} = \lvert Y \rvert e^{j\phi}$
$\lvert Y \rvert$	Magnitude of Y	siemens, S	$\lvert Y \rvert = \sqrt{G_p^2 + B_p^2} = \frac{1}{\lvert Z \rvert}$
G_p	Real part of Z	siemens, S	$G_p = \frac{R_s}{R_s^2 + X_s^2}$
B_p	Susceptance	siemens, S	$B_p = \frac{X_s}{R_s^2 + X_s^2}$

(*Contd.*)

Table 5.1 Impedance and associated terms (*Contd.*)

Parameter	*Quantity*	*Unit, symbol*	*Formula*
C_s	Series capacitance	farad, F	$C_s = \dfrac{1}{\omega X_s} = C_p(1 + D^2)$
C_p	Parallel capacitance	farad, F	$C_p = \dfrac{B}{\omega} = \dfrac{C_s}{1 + D^2}$
L_s	Series inductance	henry, H	$L_s = \dfrac{X}{\omega} = L_p \dfrac{Q^2}{1 + Q^2}$
L_p	Parallel inductance	henry, H	$L_p = \dfrac{1}{\omega B_p} = L_s\left(1 + \dfrac{1}{Q^2}\right)$
R_p	Parallel resistance	ohm, Ω	$R_p = \dfrac{1}{G_p} = R_s(1 + Q^2)$
Q	Quality factor	None	$Q = \dfrac{1}{D} = \dfrac{X_s}{R_s} = \dfrac{G_p}{B_p} = \tan\theta$
D, DF, $\tan\delta$	Dissipation factor	None	$D = \dfrac{1}{Q} = \dfrac{R_s}{X_s} = \dfrac{B_p}{G_p} = \tan(90° - \theta)$ $= \tan\delta$
θ	Phase angle of Z	degree or radian	$\theta = -\phi$
ϕ	Phase angle of Y	degree or radian	$\phi = -\theta$

Figure 5.19 shows the basic ac bridge circuit. All impedances are complex quantities. Let us denote complex quantities by overbar letters. The voltage across the detector under the balance condition should be zero. Therefore,

$$\frac{\bar{Z}_1}{\bar{Z}_1 + \bar{Z}_3}\bar{V}_S = \frac{\bar{Z}_2}{\bar{Z}_2 + \bar{Z}_4}\bar{V}_S$$

or

$$\bar{Z}_1\,\bar{Z}_4 = \bar{Z}_2\,\bar{Z}_3$$

Splitting this equation in magnitude and phase angle, we get the following two equations to be simultaneously satisfied for the balance condition:

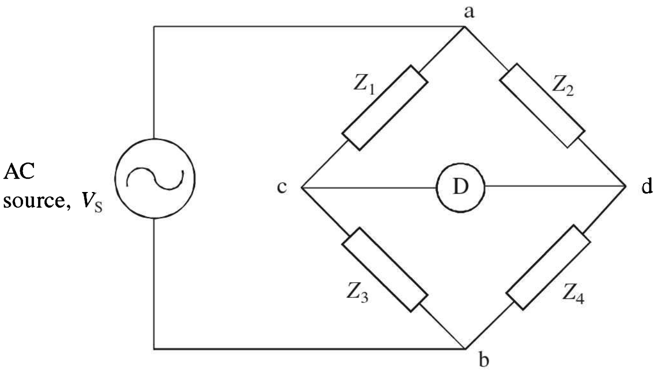


Figure 5.19 Basic ac bridge circuit.

$$|Z_1||Z_4| = |Z_2||Z_3|$$

and

$$\theta_1 + \theta_4 = \theta_2 + \theta_3$$

Similarly, for the case of admittances, we have

$$|Y_1||Y_4| = |Y_2||Y_3|$$

and

$$\phi_1 + \phi_4 = \phi_2 + \phi_3$$

The product of the magnitudes of the impedances/admittances of the opposite arms must be equal and the sum of the phase angles of the opposite arms must be equal.

Example 5.8

An ac bridge is driven by a 500 Hz sinusoidal source. Is the bridge balanced if at this frequency the bridge arms have the following impedances?

(a) $\overline{Z}_1 = 4 + j4,\quad \overline{Z}_2 = 1 + j1,\quad \overline{Z}_3 = 4,\quad \overline{Z}_4 = 1$

(b) $\overline{Z}_1 = 4 + j4,\quad \overline{Z}_2 = 1 - j1,\quad \overline{Z}_3 = 4,\quad \overline{Z}_4 = 1$

Solution

(a) Since $|Z_1||Z_4| = |Z_2||Z_3|$, we get

$$\sqrt{4^2+4^2}\sqrt{1^2+0^2} = \sqrt{1^2+1^2}\sqrt{4^2+0^2}$$

or

$$4\sqrt{2} = 4\sqrt{2}$$

This equation is satisfied.
Also

$$\theta_1 + \theta_4 = \theta_2 + \theta_3$$

or

$$45° + 0° = 45° + 0°$$

This equation is also satisfied. Hence the bridge is balanced.

(b) Since $|Z_1||Z_4| = |Z_2||Z_3|$, we get

$$\sqrt{4^2+4^2}\sqrt{1^2+0^2} = \sqrt{1^2+(-1)^2}\sqrt{4^2+0^2}$$

or

$$4\sqrt{2} = 4\sqrt{2}$$

This equation is satisfied.
Also

$$\theta_1 + \theta_4 = \theta_2 + \theta_3$$

or

$$45° + 0° = -45° + 0°$$

This equation is not satisfied. Hence the bridge is not balanced.

Example 5.9

An ac bridge is driven by a 2 kHz sinusoidal source. Given $\bar{Z}_1 = 10\ \text{k}\Omega$, $\bar{Z}_2 = 50\ \text{k}\Omega$, $\bar{Z}_3$ = series combination of a resistor of value 100 kΩ and a capacitor of value 100 μF, $\bar{Z}_4$ = series combination of an unknown resistor and an unknown reactive element. Find the value of the unknown resistor and the type and the value of the unknown reactive element.

Solution

Let the unknown resistance be denoted by R_x and the unknown reactive element be an inductor of value L_x. The balance equation of the bridge is

$$R_1(R_x + jX_{Lx}) = R_2(R_3 - jX_{C3})$$

Equating real and imaginary parts, we get

$$R_1R_x = R_2R_3 \qquad \text{(independent of signal frequency and voltage)}$$

and

$$jR_1X_{Lx} = -jR_2X_{C3}$$

This is only possible if the unknown element is also a capacitor. Therefore, we assume the unknown reactive element to be a capacitor of value C_x. This equation yields

$$-jR_1\frac{1}{\omega C_x} = -jR_2\frac{1}{\omega C_3}$$

or

$$R_1C_3 = R_2C_x \qquad \text{(independent of signal frequency and voltage)}$$

Therefore the unknown component values are

$$R_x = \frac{R_2}{R_1} R_3 = 500\ \text{k}\Omega \quad \text{and} \quad C_x = \frac{R_1}{R_2} C_3 = 20\ \mu\text{F}$$

5.3.1 Maxwell Bridge

This bridge is used to determine the value of an inductor and its quality factor in terms of known capacitance at power and audio frequencies. Figure 5.20 shows the circuit.

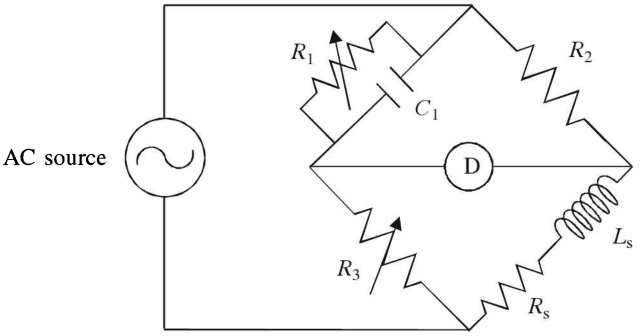


Figure 5.20 Maxwell bridge.

The balance equation is

$$\bar{Z}_1 \bar{Z}_4 = \bar{Z}_2 \bar{Z}_3$$

or

$$\bar{Z}_4 = \bar{Z}_2 \bar{Z}_3 \bar{Y}_1$$

From the bridge, we have

$$Y_1 = \frac{1}{R_1} + j\omega C_1, \quad Z_2 = R_2, \quad Z_3 = R_3, \quad Z_4 = R_s + j\omega L_s$$

Substituting in the balance equation and equating real and imaginary parts, we have

$$R_s = \frac{R_2}{R_1} R_3 \quad \text{and} \quad L_s = R_2 R_3 C_1$$

Both the equations are independent of the frequency and voltage of the signal.

The bridge is balanced first by varying R_3 for inductive balance and then R_1 for resistive balance.

Quality factor of the inductor is equal to

$$Q = \frac{X_s}{R_s} = \frac{\omega L_s}{R_s} = \omega R_1 C_1$$

This can be measured in terms of the known values of signal frequency, the value of R_1 and C_1 or from the measured values of L_s and R_s.

From the balance equation of phase angles

$$\theta_1 + \theta_4 = \theta_2 + \theta_3$$

we observe that since Z_2 and Z_3 are resistors, the right-hand side of this equation is zero (θ_2 and θ_3 are 0°). For balance, the left-hand side should also be zero. Note that θ_1 is negative (capacitive) and θ_4 is positive (inductive).

For large value of Q, R_s should be minimum, i.e. θ_4 is closer to +90°. Therefore, θ_1 should also be closer to −90°, i.e. R_1 should be very large. There may be a practical problem to get variable resistors of very large values. Maxwell bridge is, therefore, used to measure medium Q coils (inductors) say between 1 and 10.

The bridge is also not suitable for measurement of very low Q values ($Q < 1$) because of balance of convergence problem. For low value of Q, R_s is large and L_s is small, e.g. inductive resistors or *RF* coils at low frequency. Since in the equations of R_s and L_s, R_3 appears in both the equations, if for inductive balance, R_3 is decreased, it changes balance also for R_s. This is called *sliding balance*. To balance for R_s again, R_1 has to be changed. This iteration may take very long time giving rise to balance of convergence problem.

For medium values of Q, resistive effect is not large and the balance is achieved after a few iterations.

As seen from the balance equations, we may use variable C_1 instead of variable R_3 for inductive balance to avoid balance of convergence problem. But it is difficult to get accurate variable standard capacitors.

Hay's bridge is used to measure inductance of high Q values.

Example 5.10

Find the series inductance, resistance and Q of an inductive element using Maxwell bridge. Given $\bar{Z}_1$ = parallel combination of a resistor of value 470 Ω and a capacitor of value 0.22 μF, $\bar{Z}_2 = \bar{Z}_3 = 1\ \text{k}\Omega$. The bridge is driven by a 2 kHz sinusoidal source.

Solution

$$R_s = \frac{R_2}{R_1} R_3 = \frac{1000}{470} \times 1000 = 2.13\ \text{k}\Omega$$

$$L_s = R_2 R_3 C_1 = 1000 \times 1000 \times 0.22 \times 10^{-6} = 0.22\ \text{H}$$

$$Q = \omega R_1 C_1 = 2\pi f R_1 C_1 = 2 \times 3.143 \times 2000 \times 470 \times 0.22 \times 10^{-6}$$

$$= 1.3$$

5.3.2 Hay's Bridge

Hay's bridge is similar to Maxwell bridge except now R_1 is in series with C_1. This is used to measure inductances of high Q value. Figure 5.21 shows the circuit.

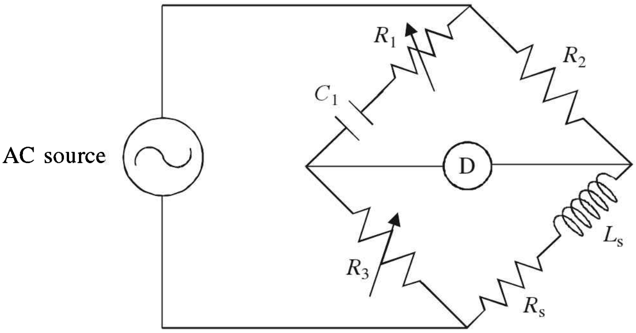


Figure 5.21 Hay's bridge.

From the bridge, we have

$$Z_1 = R_1 - \frac{j}{\omega C_1}, \quad Z_2 = R_2, \quad Z_3 = R_3, \quad Z_4 = R_s + j\omega L_s$$

Substituting in the balance equation and equating real and imaginary parts, we have

$$R_s = \frac{\omega^2 C_1^2 R_1 R_2 R_3}{1 + \omega^2 C_1^2 R_1^2} \quad \text{and} \quad L_s = \frac{R_2 R_3 C_1}{1 + \omega^2 C_1^2 R_1^2}$$

Both the equations are dependent on the frequency of the signal.

Quality factor of the inductor is equal to

$$Q = \frac{X_s}{R_s} = \frac{\omega L_s}{R_s} = \frac{1}{\omega C_1 R_1}$$

This can be measured in terms of known values of signal frequency, the value of R_1 and C_1 or from the measured values of L_s and R_s.

We can write expressions for R_s and L_s in terms of Q as follows:

$$R_s = \frac{\frac{R_2 R_3}{R_1}}{Q^2 + 1} \quad \text{and} \quad L_s = \frac{C_1 R_2 R_3}{1 + \frac{1}{Q^2}}$$

For large Q,

$$R_s \cong \frac{R_2 R_3}{Q^2 R_1} \quad \text{and} \quad L_s \cong R_2 R_3 C_1$$

We see that for high Q, the value of L_s is independent of the signal frequency and has the same relation as of Maxwell bridge. Further, the value of R_s is small compared to the relation of Maxwell bridge because of high Q.

5.3.3 Schering Bridge

This bridge is the most popular bridge and is used to measure capacitance and its insulating properties. Figure 5.22 shows the bridge circuit.

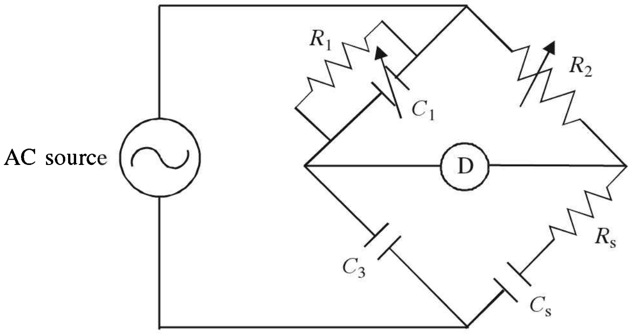


Figure 5.22 Schering bridge.

C_1 is a high quality standard capacitor. This is usually a mica capacitor, which has very low losses for general measurements, and air capacitor for insulation measurement. Air capacitor can be designed to have stable capacitance value of small electric field so that the insulating material is kept out of strong fields.

From the bridge, we have

$$Y_1 = \frac{1}{R_1} + j\omega C_1, \quad Z_2 = R_2, \quad Z_3 = -\frac{j}{\omega C_3}, \quad Z_4 = R_s - \frac{j}{\omega C_s}$$

Substituting in the balance equation and equating real and imaginary parts, we have

$$R_s = \frac{C_1 R_2}{C_3} \quad \text{and} \quad C_s = \frac{C_3 R_1}{R_2}$$

Both the equations are independent of the frequency and voltage of the signal.

Dissipation factor of the capacitor is equal to

$$D = \frac{1}{Q} = \frac{R_s}{X_s} = \omega R_s C_s = \omega R_1 C_1$$

This can be measured in terms of the known values of signal frequency, the value of R_1 and C_1 or from the measured values of C_s and R_s.

For nearly pure capacitor, D should be very low, i.e. the value of R_s should be very less which implies that the phase angle, should be nearly equal to −90°.

From the balance equation of phase angles

$$\theta_1 + \theta_4 = \theta_2 + \theta_3$$

We observe that since Z_2 is a resistor and Z_3 is a capacitor, the right-hand side of this equation is very close to –90°. For balance, the left-hand side should also be –90°. Note that θ_4 is also close to –90° for low values of D. To match the angles, Z_1 is made capacitive and C_1 is varied. In fact, at a given frequency, keeping R_1 fixed, we can directly calibrate variation of C_1 in terms of D.

Example 5.11

The Schering bridge balances under the following conditions: R_1 = 10 kΩ, R_2 = 1 kΩ, C_1 = 100 pF, and C_3 = 400 pF. The bridge is driven by a 1 kHz sinusoidal source.

(a) Find the values of R_s, C_s and D.
(b) If the series combination of R_s and C_s is to be replaced by a parallel combination R_p and C_p, find their values.

Solution

(a)

$$R_s = \frac{C_1R_2}{C_3} = \frac{100 \times 10^{-12} \times 10^3}{400 \times 10^{-12}} = 250\ \Omega$$

$$C_s = \frac{C_3R_1}{R_2} = \frac{400 \times 10^{-12} \times 10 \times 10^3}{10^3} = 4000\ \text{pF}$$

$$D = \omega R_1C_1 = 2\pi fR_1C_1 = 2 \times 3.143 \times 1000 \times 10 \times 10^3 \times 100 \times 10^{-12}$$

$$= 0.006286$$

(b) From Table 5.1, $R_p = R_s(1 + Q^2)$ and $C_p = \dfrac{C_s}{1+D^2} \cong C_s$

Now

$$Q = \frac{1}{D} = 159.08$$

Therefore,

$$R_p = 250(1 + 159.08^2) = 6.33\ \text{M}\Omega$$

and

$$C_p = 4000\ \text{pF}$$

So we can replace arm 4 with an equivalent circuit having these values.

Example 5.12

The Schering bridge is used to measure the permittivity of an insulator sample. The arm 4 consists of two electrodes in which sample is inserted. First the bridge is balanced without sample between the electrodes with the following conditions: $R_1 = R_2$ = 10 kΩ, C_1 = 100 pF and C_3 = 200 pF. Then the sample is inserted between the electrodes, and balance is achieved with the following conditions: $R_1 = R_2 = 10$ kΩ, $C_1 = 800$ pF and $C_3 = 1200$ pF. The bridge is driven by a 1 kHz sinusoidal source in each case. Find the relative permittivity of the specimen.

Solution

Let the value of the unknown capacitance without the sample be denoted by C_s and that with the sample by C_{ss}. Then

$$C_s = \frac{\varepsilon_0 A}{d} \quad \text{and} \quad C_{ss} = \frac{\varepsilon_r \varepsilon_0 A}{d}$$

where ε_r is the permittivity of the specimen, ε_0 is the permittivity of the air, A is the area of the electrodes, and d is the spacing between the electrodes. Dividing these two equations, we get

$$\text{Relative permittivity of the specimen, } \varepsilon_r = \frac{C_{ss}}{C_s}$$

From the balance equation of the bridge, we have

$$C_s = \frac{C_3 R_1}{R_2} = \frac{200 \times 10^{-12} \times 10 \times 10^3}{10 \times 10^3} = 200 \text{ pF}$$

and

$$C_{ss} = \frac{C_3 R_1}{R_2} = \frac{1200 \times 10^{-12} \times 10 \times 10^3}{10 \times 10^3} = 1200 \text{ pF}$$

Therefore, relative permittivity of the specimen, $\varepsilon_r = \frac{C_{ss}}{C_s} = \frac{1200}{200} = 6$

5.3.4 Wien Bridge

This bridge is used in many applications like notch filter, oscillators, etc. Here we shall study it as a frequency determining bridge. Figure 5.23 shows the circuit of the bridge.

From the bridge, we have

$$Z_1 = R_1 - \frac{j}{\omega C_1}, \quad Z_2 = R_2, \quad Y_3 = \frac{1}{R_3} + j\omega C_3, \quad Z_4 = R_4$$

Substituting in the balance equation and equating real and imaginary parts, we have

$$\frac{R_2}{R_4} = \frac{R_1}{R_3} + \frac{C_3}{C_1}$$

and

$$\omega^2 = \frac{1}{R_1 R_3 C_1 C_3} \quad \text{or} \quad f = \frac{1}{2\pi\sqrt{R_1 R_3 C_1 C_3}}$$

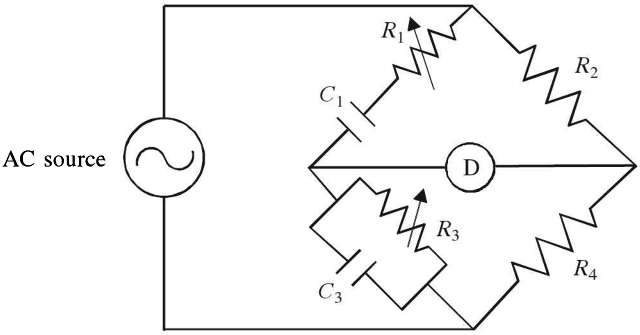


Figure 5.23 Wien bridge circuit.

In most of the bridges, components are selected so that $R_1 = R_3 = R$ and $C_1 = C_3 = C$. Then these two equations reduce to

$$\frac{R_2}{R_4} = 2 \quad \text{and} \quad f = \frac{1}{2\pi RC}$$

Keeping $R_2 = 2R_4$ and fixing $C_1 = C_3 = C$, we can calibrate $R_1 = R_3 = R$ in terms of frequency. In fact, R_1 and R_3 can be ganged together so that variation in both the resistors is same. Since the balancing of the bridge depends upon the frequency, it may be difficult to balance if the signal contains harmonics.

5.3.5 Wagner Ground Connection

The basic ac bridge circuit as shown in Figure 5.19 assumes that all the impedances are lumped and do not interact with each other. In practice, stray capacitances exist between bridge elements and the ground and across each bridge arm. These stray capacitances shunt the arms and cause error in measurements especially at high frequencies when the impedances of these capacitors become small. One way to control stray capacitances is to use shielded wires with shields connected to ground. This does not remove the effect of capacitances but makes them constant in value and thereby compensates their effect.

Wagner ground connection can be used especially to eliminate the stray capacitance effect. Figure 5.24 shows the connections. Let C_1, C_2, C_3 and C_4 be the stray capacitances between point a, d, b and c and ground respectively. Z_5 and Z_6 are impedances of the Wagner earth branch and consist of variable resistor and (or) variable capacitor. These are chosen in such a way that they can form a balance with Z_1 and Z_3 and Z_2 and Z_4.

The following procedure is adopted to balance the two bridges: Keep switch at position 1 and try to balance the bridge by varying Z_1 and (or) Z_3. The presence of stray capacitances will prevent a true balance but a point of minimum detector reading can be achieved. Change the position of the switch to 2 to balance the bridge by varying Z_5 and (or) Z_6 to a point of minimum detector reading. Again, change the position of the switch to position 1 and rebalance the bridge by varying Z_1 and (or) Z_3. Repeat the process till the detector reads zero both for positions 1 and 2 of the switch. Under this condition, points c and d are at ground

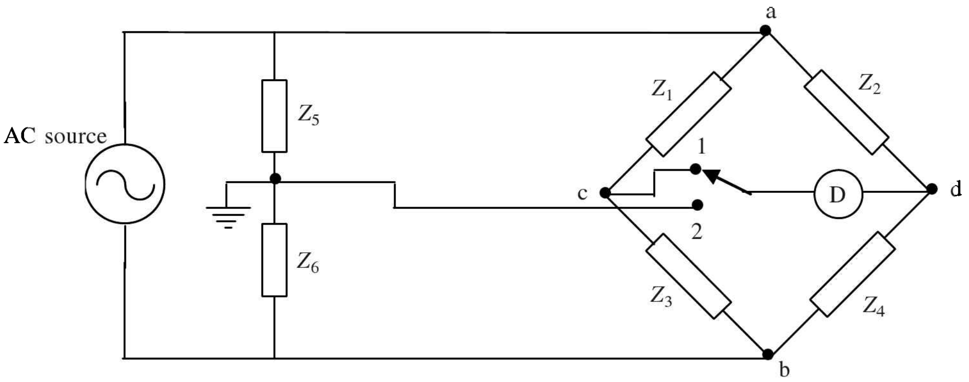


Figure 5.24 Wagner ground connection.

potential and the effect of stray capacitances C_2 and C_4 is nullified. Similarly, capacitors C_1 and C_3 are connected between points a and b and ground, i.e. across Z_5 and Z_6 respectively which do not form the arms of the original bridge Z_1, Z_2, Z_3 and Z_4.

The main disadvantage of the Wagner ground connection is that the two balances must be made for each measurement—one to balance the bridge, and the other to balance the Wagner ground.

5.4 ANDERSON LOOP

Wheatstone bridge provides a number of well-known advantages but has many disadvantages in critical applications. Some of these disadvantages are:

- Half the signal from each element's impedance change is typically consumed by adjacent bridge arms.
- The output signal is usually a nonlinear function of impedance change per individual bridge arm.
- Each individual sensing impedance variation is typically able to influence only one measurement output.
- Lead wire and connector impedance changes typically add measurement uncertainty, particularly when they occur in the inter-bridge wiring and especially when the various changes do not evolve identically.
- As many as eight lead wires may be required per bridge transducer to impose a reliable excitation level and to precisely alter the output of the transducer to accomplish an end-to-end system electrical calibration.
- Multiple sensing impedances observed simultaneously provide only a single output signal that is typically a fixed function of the sensor element impedance changes.
- The transfer function, including compensation, of a bridge transducer is essentially *locked in*—typically not adjustable after manufacture and very rarely adjustable after installation when the actual operating environment becomes known.
- Thermoelectric (self-generating) outputs are difficult to separate from continuous impedance change (non-self-generating) outputs when using dc excitation.

Anderson loop measurement circuit topology deals effectively with all of these limitations by using *active* subtraction to observe the change in the output level of the variable impedance element(s) instead of the *passive* subtraction accomplished by the Wheatstone bridge circuit.

The benefits of employing active subtractors in measurement applications are enormous. Many things previously considered impossible become routine, for example, the continuous separation of thermoelectric signals from IR drop signals. A strain gauge can be connected in an Anderson loop with thermocouple wire and, with dc excitation, strain and temperature outputs are available, each essentially uninfluenced by the other.

The Anderson loop circuit topology is universal in that, resistances, inductances and capacitances can be used as gauge and reference elements with dynamic loop excitation waveforms. Figure 5.25 shows the basic dual-differential subtractor (Anderson subtractor).

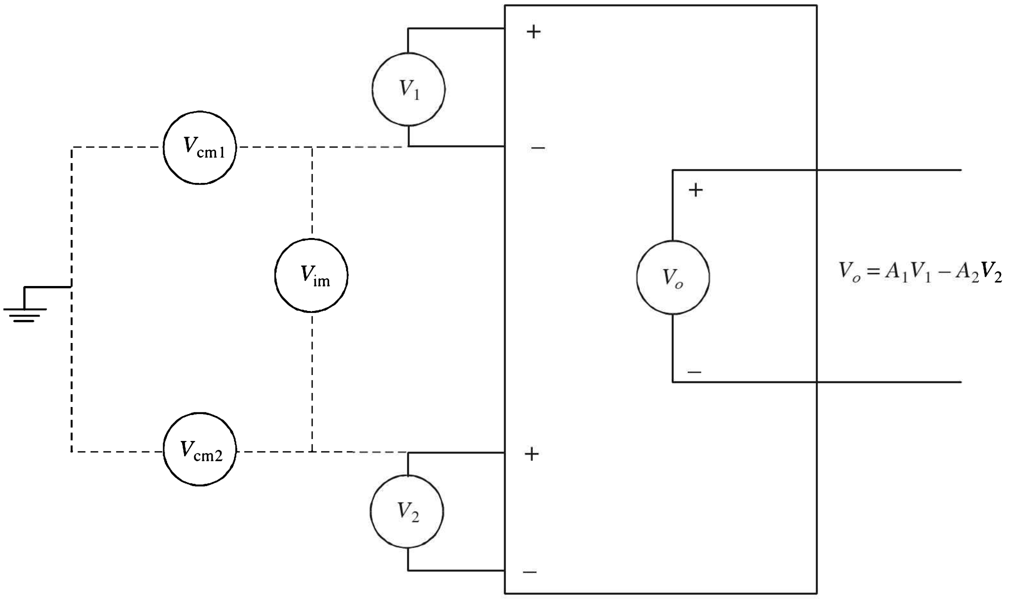


Figure 5.25 Dual-differential subtractor (Anderson subtractor).

The ideal dual-differential subtractor is a six-terminal three-port active circuit that delivers at its output, V_o, equal to the difference between the two input potential differences, V_1 and V_2, amplified by gains A_1 and A_2, respectively, and is uninfluenced by any common potential difference, V_{cm1} and V_{cm2}, or difference (interior) mode potential difference from one input to the other V_{im}.

Anderson loop measurement circuit topology

Figure 5.26 shows the Anderson loop measurement circuit topology.

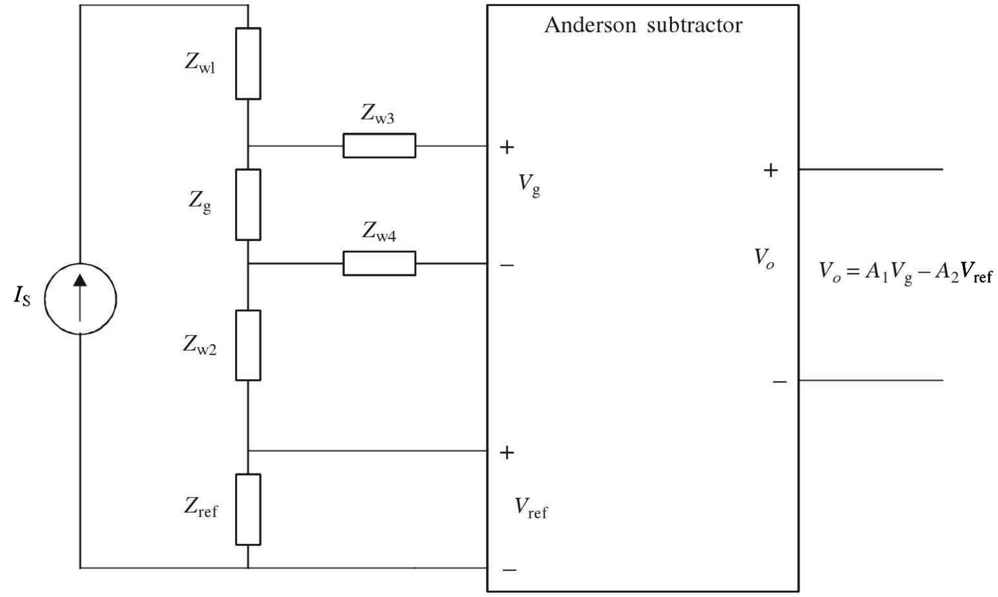


Figure 5.26 Anderson loop measurement circuit topology.

I_S is a constant current source and is equal to V_{ref}/Z_{ref}. Z_g is the unknown impedance. Z_{w1}, Z_{w2}, Z_{w3}, and Z_{w4} are the wiring impedances. Z_{w1} and Z_{w2} do not affect the voltage V_g. Also since the input impedance of the Anderson subtractor is large, no significant voltage drop occurs across the sensing wires with impedances Z_{w3} and Z_{w4}. The output voltage of the subtractor is given by

$$V_o = I_S\,[A_1(Z_g + \Delta Z) - A_2Z_{ref}]$$

where ΔZ is the change in Z_g.

When $A_1 = A_2$ and $Z_g = Z_{ref} = Z$, we get

$$V_o = I_S\Delta Z$$

The result indicates that the loop has an inherently linear response to changes in a sensor element and no impedance terms are present to attenuate the output. Also, unlike typical Kelvin observations, any variation in the excitation level, I_S, appears in the output as a per cent-of-reading error rather than as a per cent-of-excitation error. Further, compared to Wheatstone bridge output (Eq. 5.2), the output in Anderson Loop is very large. Using only 25% of the excitation power, the Anderson loop can deliver the same sensor output voltage level as a Wheatstone bridge. This tremendous increase in efficiency is because the Anderson loop output voltage is double of what can be obtained from a Wheatstone bridge having the same excitation voltage across each of the various sensor elements.

We can express this equation in terms of the reference voltage

$$V_o = \frac{\Delta Z}{Z}V_{ref}$$

In terms of ratiometeric observations,

$$\frac{\Delta Z}{Z} = \frac{V_o}{V_{\text{ref}}}$$

Ratiometeric operation of the Anderson loop is accomplished by simply using V_{ref}, the voltage drop across the reference element as the reference input to the system observing V_o, for example, an A-to-D converter. With ratiometeric operation, regulation of the excitation current becomes theoretically unnecessary.

Multiple sensing elements can be observed with respect to one or more reference elements by using multiple subtractors in the same Anderson loop. For example, changes in the four resistance elements in a typical Wheatstone bridge transducer can be independently observed when they are wired in an Anderson loop. Small differences developing between widely separated sensor elements can be reliably observed.

Large variations in the various lead wires do not significantly affect practical Anderson loop measurements. Connecting wires can be of any conducting material and only need to be large enough to survive in the measurement environment.

When the sensing elements are physically near each other we can save on the number of connecting wires used (see Example 5.13). For example, a strain gauge rosette (three co-located sensors sensitive to strain at various angles) can be observed with only six connecting wires. Even with only two wires per sensor, connecting wire resistance can vary randomly due to environmental effects without introducing a significant error in strain rosette measurements.

Any active element in a measurement system adds some noise to the measured result. However, the overall noise floor of an Anderson loop measurement is lower than the noise observed using Wheatstone bridge circuits as long as the noise in the measurement environment is significantly greater than the noise added by the subtractor. An improvement in signal-to-noise ratio of the order of 6 dB when compared to the Wheatstone bridge circuit can be achieved.

Let us summarize the benefits of the Anderson loop compared to Wheatstone bridge.

- Only 25% of the excitation power provides the same output level from a transducer.
- The output is double of what is available from the Wheatstone bridge for the same power dissipation (excitation level) in a transducer. Strain gauges appear to double in the gauge factor.
- Each lead wire resistance can vary randomly; lead wire resistance becomes irrelevant in a measurement application, a major benefit in severe environments.
- The impedance elements of a bridge circuit can be simultaneously observed independently and in any arithmetic combination.
- Resistance variations and temperature can be independently and continuously monitored by connecting to a sensor with thermocouple wires.
- The output is a linear function of sensor resistance or impedance change—even for changes in a single element of what was previously wired as a bridge-topology circuit. This means that single strain gauge measurements are inherently linear when using Anderson loop signal conditioning.
- Transducers can withstand twice the overload for the same sensitivity.

- Transducers can have twice the upper frequency response for the same sensitivity.
- Transducers can have both factory- and user-variable compensation.
- Fewer wires can be used to connect to the transducer; a strain rosette requires only six wires yet random wire resistance changes are irrelevant with only two wires per rosette gauge.
- Transmitters are not required to avoid lead-wire resistance change and thermoelectric effects.
- Smart transducers can have different calibrations applied to each individual element they contain.
- Resistance and impedance differences between widely-separated sensor elements are obtained with outstanding, absolute and differential accuracy, yielding a dispersed sensor.

Example 5.13

How will you use Anderson loop for measurement of microstrain of a resistive strain gauge with gauge factor, *G*?

Solution

For a resistive strain gauge, the strain is given by

$$\varepsilon = \frac{\frac{\Delta R}{R}}{G} = \frac{V_o}{V_{\text{ref}} G}$$

or microstrain is given by

$$\mu \text{ strain} = \frac{10^6 V_o}{V_{\text{ref}} G}$$

If V_o is amplified by A, then

$$\mu \text{ strain} = \frac{10^6 V_o}{A V_{\text{ref}} G}$$

So, knowing V_o, V_{ref}, G and A, we can find microstrain of the gauge.

Example 5.14

Figure 5.27 shows the use of instrumentation amplifiers (IA) as half-Anderson subtractors. Each IA acts as one-half of a dual-differential subtractor by observing sensor and reference voltage drops across several sensing elements in a loop: When the initial impedance of each sensing element is the same as the reference impedance and the gain of each IA is unity, find all the output voltages of the topology.

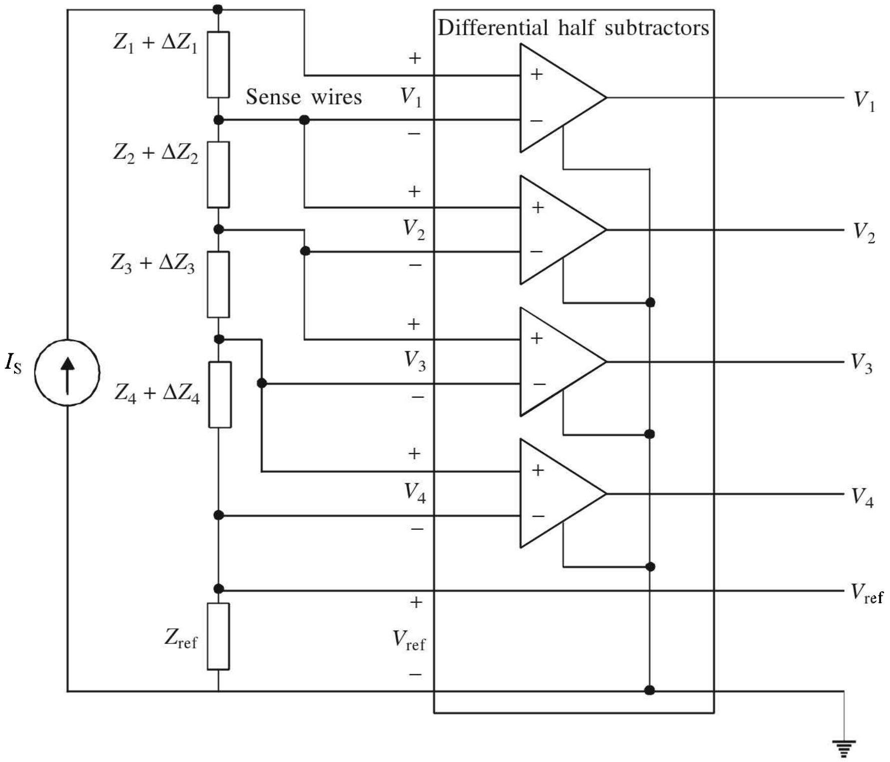


Figure 5.27 Multi-sensor loops using Anderson half subtractors.

Solution

The output voltage of the IA with Z_4 input is given by

$$V_4 - V_{ref} = I_S[(Z_4 + \Delta Z_4) - Z_{ref}] = I_S\,\Delta Z_4 \qquad \text{(since } Z_4 = Z_{ref}\text{)}$$

or

$$V_4 = I_S\,\Delta Z_4 + V_{ref} \tag{i}$$

The output voltage of the IA with Z_3 input is given by

$$V_3 = I_S\,\Delta Z_3 + V_{ref} \tag{ii}$$

Similarly, the output voltages of the IAs with Z_2 and Z_1 inputs are given by

$$V_2 = I_S\,\Delta Z_2 + V_{ref} \tag{iii}$$

$$V_1 = I_S\,\Delta Z_1 + V_{ref} \tag{iv}$$

By subtracting these outputs, we can get the relative outputs independent of V_{ref} represented by the following equations:

$$V_1 - V_2 = I_S(\Delta Z_1 - \Delta Z_2)$$

$$V_1 - V_3 = I_S(\Delta Z_1 - \Delta Z_3)$$

$$V_1 - V_4 = I_S(\Delta Z_1 - \Delta Z_4)$$

$$V_2 - V_3 = I_S(\Delta Z_2 - \Delta Z_3)$$

$$V_2 - V_4 = I_S(\Delta Z_2 - \Delta Z_4)$$

$$V_3 - V_4 = I_S(\Delta Z_3 - \Delta Z_4)$$

This wiring configuration shares sense lead-wires. If the negative end of one sensor impedance, say Z_1, and the positive end of an adjacent impedance, say Z_2, are both at essentially the same potential, then the same sense lead-wire can accomplish both observations. This configuration requires a total number of lead-wires equal to three plus the number of impedances in the loop.

5.5 LCR BRIDGE

An LCR bridge is an instrument used to measure impedance and, therefore, all parameters as mentioned in Table 5.1. This is also called *Universal Bridge* and contains several types of resistance, capacitance and inductance measuring bridges. The bridge is chosen manually depending upon the Device Under Test (DUT). Since iterative balancing manipulation is needed in these bridges to reach the final null (balance) condition, many automatic bridges have been developed in order to eliminate time consuming manual operation and associated measurement errors. These are also called *Digibridges* or *LCR meters.*

Most recent instruments employ microprocessors, elaborate software driven control and signal processing techniques. They use principle of measurement which differs significantly from that employed by the traditional measuring instruments. Figure 5.28 shows an elementary measurement circuit.

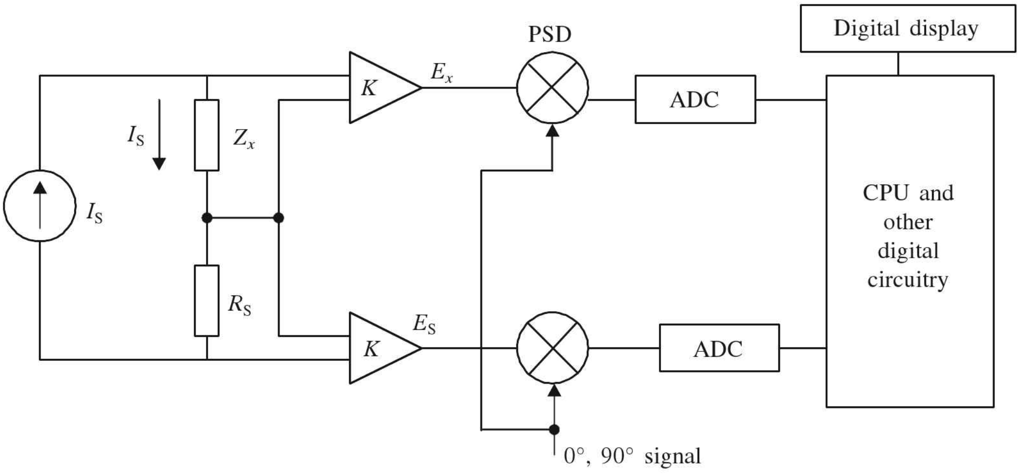


Figure 5.28 Automatic impedance bridge.

Both the voltage across the device under test impedance Z_x, and the voltage across a reference resistor R_S, are measured, which essentially carry the same current. The voltage across Z_x is E_x and the voltage across R_S is E_S, i.e.

$$E_x = KI_S Z_x$$

and

$$E_S = KI_S R_S$$

Therefore,

$$Z_x = R_S \frac{E_x}{E_S}$$

Both voltages are simultaneously sampled many times per cycle of the applied sine wave excitation. For both E_x and E_S, real and imaginary components are computed from the individual sample measurements data. The real and imaginary components of E_x (E_{xr} and E_{xi}) are combined with the real and imaginary components of E_S (E_{Sr} and E_{Si}) and the known characteristics of the reference resistor to determine the apparent impedance of the complex impedance of Z_x using complex arithmetic.

A phase sensitive detector (PSD) is used to extract only the resistance element of Z_x. The output of the PSD is converted to digital signal using analog-to-digital converter. CPU and other digital circuitry perform mathematical calculations and calculate all the desired parameters of the DUT. This circuitry and fast sampling rates, along with elaborate software driven control, can provide direct readout of all the impedance parameters previously discussed.

Four-terminal measurements

To neglect the effect of lead resistance, four-terminal measurements (as explained in Chapter 2) are provided to obtain accurate impedance measurements with today's LCR meters. One pair of leads applies the current (source) to the device while the other pair of leads measures (senses) the voltage across the device. This removes the series inductance and resistance error factor including contact resistance. It also removes the stray capacitance factor. Shielded coaxial cabling will reduce most mutual inductance that occurs between the current leads and the voltmeter leads. This four-terminal connection technique is widely known as a *Kelvin* connection.

Performance parameters

Test signal level. This specifies the test signal level applied to the DUT at measurement. This specification becomes important for DUTs whose impedance varies with the signal level. Test signal level may also be specified by the industry standards. The ac output of most LCR meters can be programmed to select the signal level applied to the DUT. Generally, the programmed level is obtained under an open-circuit condition. A source resistance internal to the meter is effectively connected in series with the ac output and there is a voltage drop across this resistor. When DUT is connected, the voltage applied to the device depends on the value of the source resistance and the impedance of the device. For example, if both are equal, then signal level applied to the DUT will be half. Some LCR meters have a constant voltage mode, where the voltage to the device is monitored and maintained at the programmed level.

Averaging. This is a way of improving instrument measurement accuracy, by reducing unwanted signals and effects of unwanted noise, but some measurement speed is sacrificed. In an averaging mode many measurements are made and the average of these calculated for the final result. With more measurements averaged, the accuracy is enhanced. The length of time that an LCR meter spends integrating analog voltages during the process of data acquisition can have an important effect on the measurement results. If integration occurs over more cycles of the test signal the measurement time will be longer, but the accuracy will be enhanced.

Autoranging. On an impedance-measuring instrument, autoranging is intended to maintain maximum signal level and the highest signal-to-noise ratio for maximum measurement accuracy. In order to measure over wide impedance ranges, an instrument must have several measurement ranges. The measured impedance should be as close as possible to the full-scale on any given range. Ranging is usually done automatically depending on the impedance of the test device, thus eliminating an improper selection by an operator resulting in erroneous measurement data of test device. Range changes are accomplished by switching range resistors and the gain of detector circuits.

Range holding. Range holding, rather than autoranging, is a feature sometimes used in specific applications. For example, when repetitive testing of similar value components is to be done, range holding can speed up the test time. Another use of range hold occurs when measuring components whose values fall within the overlap area of two adjacent ranges, where if allowed to autorange an instrument display can sometimes result in operator confusion.

Open/short compensation. It reduces the effects from error sources between the DUT and the calibrated connection to the measuring instrument. Compensation ensures the best measurement accuracy possible on a device at selected test conditions. The traditional open/short routine measures the stray admittance (Y_{OPEN}) and residual impedance (Z_{SHORT}) of the test cable(s) to remove these quantities from the actual measured DUT value. When the device is measured, these two residuals are used to calculate the actual impedance of the DUT. The OPEN routine determines the correction that the instrument applies to high-impedance measurements, and the SHORT routine determines the correction necessary for low-impedance measurements. While performing an OPEN measurement it is important to keep the distance between the unknown terminals the same as they are when attached to the device. It is equally important to make sure that we does not touch or move our hands near the terminals. When performing a SHORT measurement, a shorting device (shorting bar or highly conductive wire) is connected between the terminals. For very low-impedance measurements, it is best to connect the unknown terminals directly together.

Load correction. It is a compensation technique that uses a load whose impedance is accurately known, and applies a correction to measurements of similar components to substantially improve measurement accuracy. The purpose of this is to correct for non linearity of the measuring instrument, and test fixture or lead effects (which may be dependent on the test frequency, test voltage, impedance range, or other factors), all of which may be repeatable and thus can be added or subtracted each time a similar measurement is made. Load correction is only as good as the known accuracy and/or traceability of the load used in determining the correction.

Criteria for selecting the appropriate load include:

- Load whose impedance value is accurately known.
- Load whose impedance value is very close to the DUT (this ensures that the measuring instrument selects the same measurement range for both devices).
- Load whose impedance value is stable under the measurement conditions.
- Load whose physical properties allow it to be connected using the same leads or fixture as the DUT.

A prerequisite for load correction is to perform careful open/short compensation as previously discussed.

$$Z_{\text{actual}} = Z_{\text{measured}} \pm \Delta Z$$

where ΔZ is the difference between the known and measured value of the standard.

Through the use of load correction it is possible to effectively increase the accuracy of the measuring instrument substantially, but this is only as good as the known accuracy and traceability of the load used in determining the correction.

Computer interface. Many testers today are equipped with some type of standard data communication interface for connection to remote data processing, computer or remote control. The typical interfaces are IEEE-488 (GPIB) or the RS-232 (serial communication) line. These interfaces are commonly used for monitoring trends and process control in a component manufacturing area or in an environment where archiving data for future reference is required. The whole idea of production monitoring is to reduce yield risks and be able to correct the process quickly if needed. An LCR meter with remote interface capability has become standard in many test applications where data logging or remote control have become commonplace.

QUESTIONS

5.1 A highly sensitive galvanometer can detect a minimum current of 0.1 nA. This galvanometer is used in a Wheatstone bridge as a detector. The resistance of the galvanometer is negligible. Each arm of the bridge has a resistance of 1 kΩ. The source voltage connected to the bridge is 20 V. Calculate the smallest change in resistance which can be detected. The resistance of the galvanometer can be neglected compared to the resistances of the bridge.

5.2 Figure Q5.2 shows a single-element varying active bridge in which an opamp produces a forced null, by adding a voltage in series with the variable arm. The voltage is equal in magnitude and opposite in polarity to the incremental voltage across the varying element and linearly with ΔR. Since it is an opamp output, it can be used as a low impedance output point for the bridge measurement. This active bridge has a gain of two over the standard single-element varying bridge, and the output is linear, even for large values of ΔR. Assuming an ideal opamp, verify that the output voltage is given by

$$V_{\text{out}} = -V_B\left(\frac{\Delta R}{2R}\right)$$

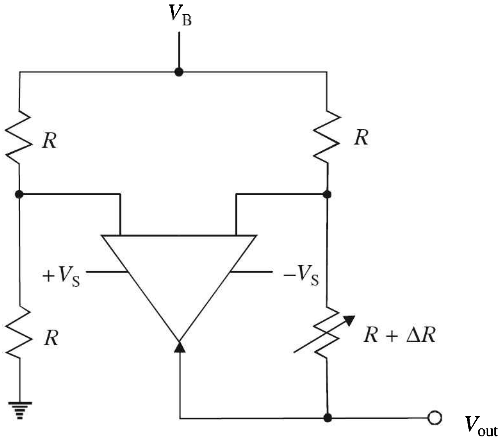


Figure Q5.2

5.3 If the bridge of the circuit of Figure Q5.2 is changed to two-element varying active bridge by varying resistance between V_B and non-inverting input as well, show that

$$V_{\text{out}} = -V_B\left(\frac{\Delta R}{R}\right)$$

5.4 Another circuit for linearizing a single-element varying bridge is shown in Figure Q5.4. The bottom of the bridge is driven by an opamp, which maintains a constant current in

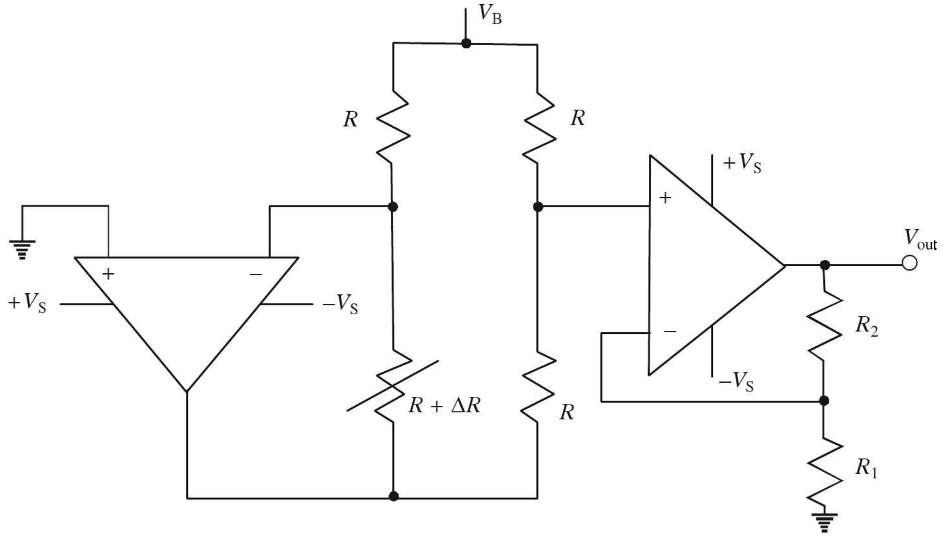


Figure Q5.4

the varying resistance element. The output signal is taken from the right-hand leg of the bridge and amplified by a non-inverting opamp. The output is linear, but the circuit requires two opamps. In addition, R_1 and R_2 must be matched for accurate gain. Assuming ideal opamps, verify that the output voltage is given by

$$V_{\text{out}} = \frac{V_B}{2}\left(\frac{\Delta R}{R}\right)\left(1 + \frac{R_2}{R_1}\right)$$

5.5 A strain bridge comprises two 150 Ω resistors, one active gauge and one unstrained gauge for temperature compensation. The two gauges have unstrained resistances of 150 Ω and a strain gauge factor of 2.5. The bridge supply voltage is 5 V. Calculate the strain when the voltmeter reading is 2 mV.

5.6 In a Wheatstone bridge, sensitivity to unbalance is measured with two galvanometers of equal sensitivity but unequal internal resistances. Which galvanometer will show more sensitivity to unbalance?

5.7 Compare the performances of a constant voltage and constant current Wheatstone bridge.

5.8 How is the effect of leakage taken care of for the measurement of high resistance with the Wheatstone bridge?

5.9 How is the variation in excitation voltage taken care of in an all element varying remotely connected Wheatstone bridge?

5.10 Which bridge will you use for the measurement of the resistance of an RF coil? Explain its operation.

5.11 In a Maxwell bridge, given $\bar{Z}_1$ is the parallel combination of a resistor of value 200 Ω and a capacitor of value 5 μF, $\bar{Z}_2$ = 0.5 kΩ, $\bar{Z}_3$ = 0.4 kΩ and $\bar{Z}_4$ is the series combination of a resistor of value 1 kΩ and an inductor of value 1 H. The bridge is operated at ω = 1000 rad/s sinusoidal source. Is the bridge balanced?

5.12 (a) The quality factor Q of a coil is ≥ 10. Suggest a bridge to measure the values of inductor and its series resistance. Express these values in terms of Q.

(b) How can you measure the unknown value of applied frequency (ω or f) using this bridge?

5.13 It is desired to measure permittivity of an insulator sample. Suggest a bridge for the same and explain the procedure.

5.14 Schering bridge is used for the measurement of permittivity of a specimen. Without specimen between plates of C_S, balance is obtained with values: C_1 = 120 pF, C_3 = 150 pF, $R_1 = R_2$ = 5 kΩ. With the specimen inserted between plates of C_S, values for balance become: C_1 = 200 pF, C_3 = 900 pF, $R_1 = R_2$ = 5 kΩ. In each test ω = 5000 rad/s. Find the relative permittivity of the specimen.

5.15 Explain Wagner ground connection.

5.16 Figure Q5.16 shows an inverted Wien bridge where R_4 and C_4 are unknown. R_1 = 1008 ± 0.5 Ω and R_3 = 2002 ± 0.5 Ω. At the balance point, R_2 = 1080 ± 5 Ω and C_2 = 190 ± 0.5 nF. Calculate R_4 and C_4. Calculate the fractional uncertainties in the values of R_4 and C_4.

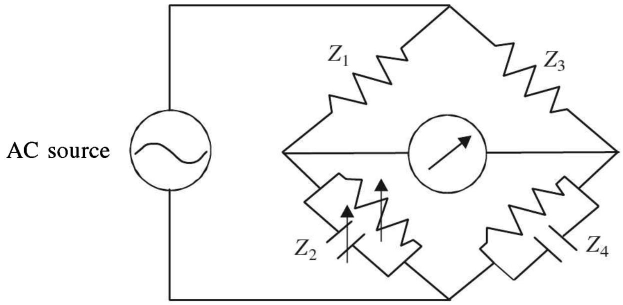


Figure Q5.16

5.17 The circuit of Figure Q5.17 shows a bridge to measure impedance of both capacitive and inductive circuits at higher frequencies. The bridge is first balanced with Z_x terminals shorted and null obtained with the following values:

$f = 10$ MHz, $R_2 = 8$ kΩ, $R_3 = 27.5$ kΩ, $R_4 = 7.5$ kΩ, $C_1 = 7.5$ μF, $C_2 = 120$ μF, $C_3 = 100$ μF

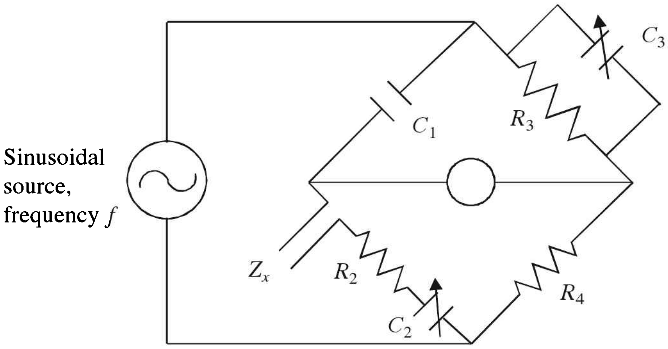


Figure Q5.17

Now the unknown impedance ($R_x + jX_x$) is inserted at Z_x terminals and rebalance is obtained by varying C_2 and C_3 to following values: $C_2 = 102.4$ μF, $C_3 = 110$ μF. Find the value of R_x and X_x. Also find the value of unknown reactive component.

5.18 Draw a typical Anderson loop measurement circuit topology and explain its operation. What are its advantages over Wheatstone bridge?

5.19 In Figure Q5.19, a Type T (copper/copper-nickel) thermocouple wire pair is used both to connect excitation I and to sense the voltage level at each terminal of a variable impedance sensor ($R + \Delta R$)—a strain gauge in the system. Isothermal block represents the "reference temperature" where the wires are connected to the remote strain gauge. R_{ref} is selected to be essentially equal to the initial resistance of the strain gauge, R. R_{w1} is essentially equal to R_{w2} and R_{w3} is equal to R_{w4}. $v_{11} - v_{12}$ and $v_{21} - v_{22}$ are defined as the net thermoelectric components, v_{TC1} and v_{TC2}, respectively. Show that $v_1 - v_3 = 2v_{TC}$ and $v_2 - v_4 = I\Delta R$. We see that the desired thermoelectric output is uninfluenced by

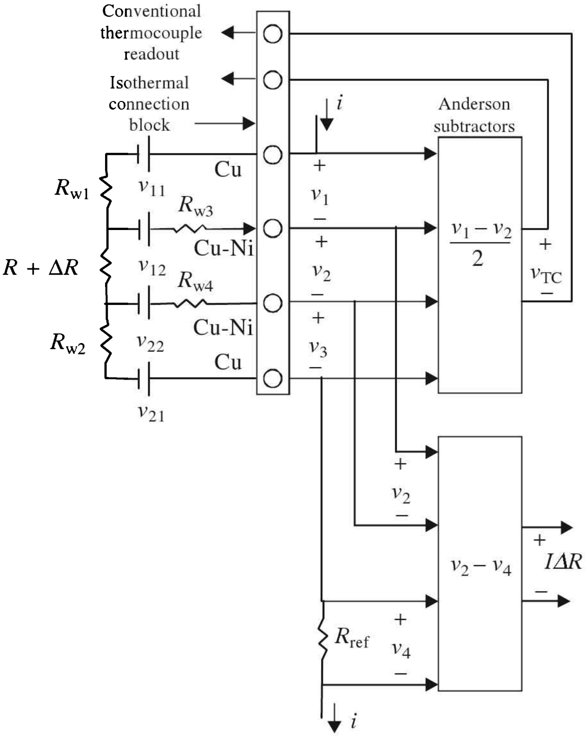


Figure Q5.19

excitation current flowing through the thermocouple wires and the desired strain gauge resistance change output as a function of excitation current and is uninfluenced by either thermoelectric voltages or random variations in wire resistance.

5.20 Draw the block diagram of an automatic impedance bridge. Explain its operation and performance parameters.

5.21 Write typical specifications of an LCR Digibridge.

Chapter 6

Recorders and Data Loggers

6.1 INTRODUCTION

There is often a need to store any measurand for current and future reference. Today, methods of data storage comprise both digital computer memory and traditional paper. Broadly speaking, there are two principal areas where recorders or data loggers are used:

(i) Process measurements for such variables as temperature, pressure, flow, pH, and humidity

(ii) Scientific and engineering applications such as high-speed testing (e.g. stress/strain), statistical analyses, and other laboratory or off-line uses where a graphic or digital record of selected variables is desired.

Instrumentation for each of these fields can have many operating principles and features in common. Each area, however, may place more emphasis on the importance of such factors as cost, method of use (e.g. bench-top vs. panel mounting), portability, degree of accuracy, speed of response, and number of points (channels) to be measured.

Digital computer systems have the ability to provide useful trend curves on CRT displays that could be analyzed and printed out. A similar approach permits digital inputs to recorders and data loggers to be stored electronically on-board, or sent to a remote PC, or distributed control system (DCS) workstation, or printer.

There is now a plethora of data acquisition hardware and software that can work with PCs, making it quite plausible to classify them as recorders, or alternatively data loggers.

There remains a continuing role for conventional standalone recorders that still use paper and pen or printer, but here too, such products have gained greatly in capabilities due to advances in electronics and new technologies such as inkjet and laser printing. New technology has also made paperless or videographic recorders a practical reality, with savings in materials and maintenance.

Traditional data loggers, whose original basic function was to print out digital values of a measured variable versus time, also have benefited from new technology that has added disk or card memory storage and faster logging capabilities. Some can provide real-time recording via network connection to a host system. In recent years, conventional analog recorders have had the functions of a data logger added to their capabilities. Called as *hybrid recorders*, they combine analog trend representations with digital information on the same chart paper.

Scores of specialized software packages are available not only to implement the PC-based functions mentioned above but also to perform complex mathematical operations or signal conditioning on input and output data.

For high-speed recording, the traditional analog oscilloscope has also been enhanced by a digital storage version (DSO), which finds an important application in test and analysis of today's many electronic products—both consumer and industrial.

6.1.1 Distinction between a Recorder and a Data Logger

There are no sharp lines distinguishing the two devices. A traditional recorder accepts an input and compares it to a full-scale value. The recording pen is then moved to a point on the scale that represents a useful value of the input, say, 500°C with a chart range of 0 to 1000°C, at a time that is given by the calibrations on the recorder's moving chart. On the other hand, a data logger accepts an input which is fed into an analog-to-digital converter and stored digitally, or printed out as a series of time-stamped values. Today, many such instruments have some form of on-board intelligence which provides the user with diverse capabilities for signal conditioning as well as data storage and analysis.

As an alternative to a recorder, a data logger can have the ability to accept a greater number of inputs (referred to as channels), with better resolution and accuracy. Only a recorder, however, can provide a truly continuous trend display of the variable's change with time.

Most of the recorders are electromechanical devices. They consist of a paper chart on which information is recorded, a writing instrument such as a stylus or a pen, and interfacing apparatus between the measured parameter and the writing instrument. They can be broadly classified as:

(a) Instruments that record one or more parameters that vary with time. Such instruments are called *strip chart recorders*. These may be self-balancing or of galvanometric type.

(b) Instruments that record one or more dependent parameters that vary with respect to some independent parameter. Such instruments are called *X–Y recorders* or function plotters.

Although recorders that use mechanical writing devices like pen or stylus have very limited frequency response (up to about 125 Hz), they are widely used in industry where this frequency response is adequate. Recorders which use photographic paper for recording, however, have frequency response of a few thousand hertzs.

6.2 STRIP CHART RECORDER

6.2.1 PMMC Galvanometer Type

In this type of recorder, the pointer of PMMC galvanometer is replaced by a pen which is attached to the moving coil in the field of a permanent magnet. Deflection of the coil pen assembly is directly proportional to the current through the coil. The tip of the pen is positioned over a strip of chart paper or Z fold paper that is pulled under the pen tip at a constant speed thereby establishing a time base. The deflection of the pen across the paper reproduces the wave shape of the signal applied to the coil.

A PMMC galvanometer with a short pen arc writes in a curvilinear manner, as shown in Figure 6.1. Instead of a straight line, the pen scribes an arc and the displacement $y = R \sin \theta$, where R is the length of the pointer and θ is the angular deflection of the coil of PMMC. This relation is nonlinear. If θ is kept less than ±10°, then the error due to this nonlinearity is less than 0.5%.

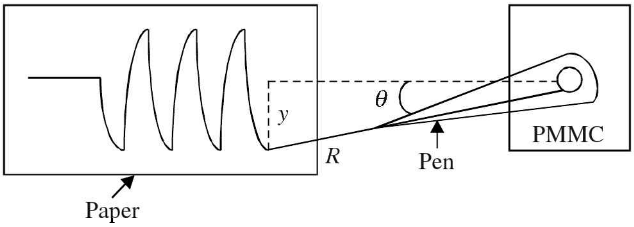


Figure 6.1 Curvilinear recording.

In many applications where the amount of deflection is more important than wave shape, this may be acceptable. One solution to the problem of curvilinear recording is to use a pivoted pen motor assembly, as shown in Figure 6.2, which translates the curvilinear motion of the moving coil to the rectilinear motion at the pen's tip. Another way is to increase the length of the pen so that it is very large compared to the width of the chart as shown in Figure 6.3. The trace then will be nearly linear and is acceptable in many applications. This is called *pseudorectilinear* technique.

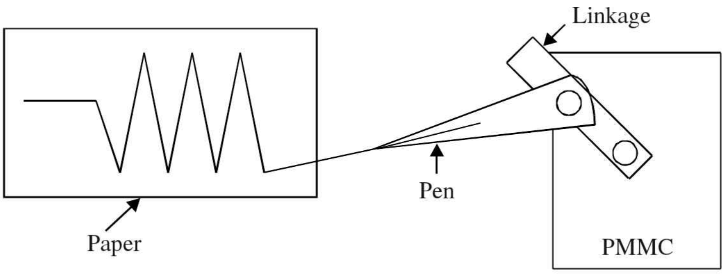


Figure 6.2 Rectilinear recording.

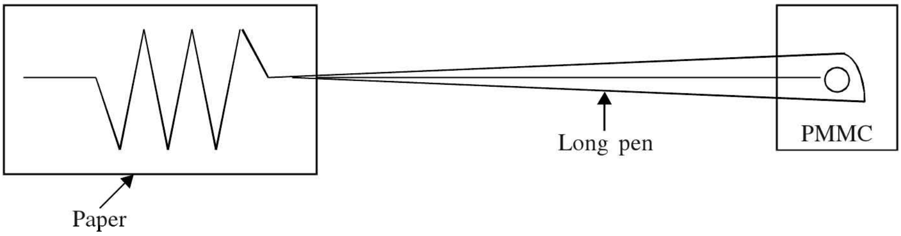


Figure 6.3 Pseudorectilinear recording.

PMMC writing mechanisms

Following writing mechanisms are used on PMMC recorders.

1. *Ink filled pen (stylus).* Felt pens or disposable fibre tipped pens can also be used.
2. *Thermal ink pen.* Here stylus is heated which blackens the treated paraffin wax paper.
3. *Ink jet.* In this mechanism, ink at high pressure is fed to a nozzle mounted on the PMMC galvanometer instead of a pen. The ink jet is directed at the paper and when the system is properly adjusted, it produces a fine line.
4. *Direct pressure (chopper bar).* This mechanism uses a special treated paper with carbonized back. The meter pointer taps the point where a mark is required. A black mark will appear on the paper at the pressure point.
5. *Optical mechanism.* It uses a mirror mounted on the galvanometer to reflect a light (generally ultra violet) beam from a collimated light source onto the photosensitive chart paper. See Figure 6.4. The paper is exposed to ultraviolet light that develops the trace as paper comes out of the recorder. This paper should be stored in a light-tight box, otherwise it will fade out with time.

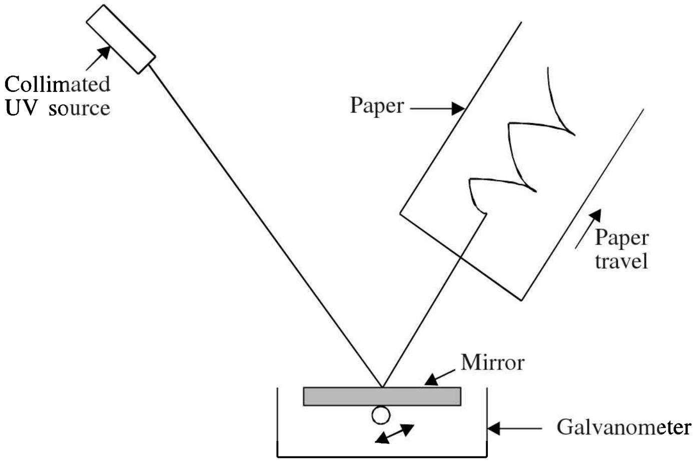


Figure 6.4 Galvanometric type optical recorder.

Optical recording is also used in CRT optical recorders. See Figure 6.5. There is no time base sweep on the CRT screen. The beam sweeps back and forth in the Y-direction only in response to amplitude variation of the input signal. Time base is provided by the speed at which paper moves across the screen.

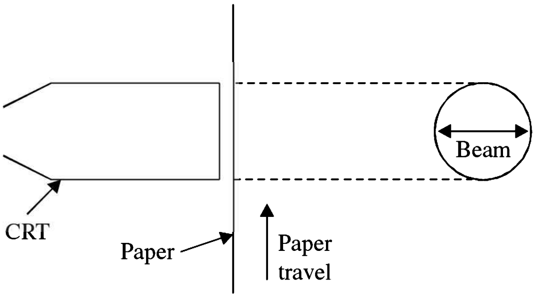


Figure 6.5 CRT camera optical recorder.

Those mechanisms that use pen or stylus have relatively low frequency response (100 to 200 Hz) due to the inertia of pen and stylus. On the other hand mechanisms like ink jet and optical have higher frequency response (1 kHz to 2 kHz). CRT camera has the highest frequency response because it is limited only by the photographic writing speed of the recording paper.

Dynamic behaviour of galvanometric recorders

The deflecting torque, T_D, of the PMMC movement is given by

$$T_D = BANI \tag{6.1}$$

where B is flux density in the air gap, A is the effective coil area, N is the number of turns of wire on the coil and I is the current in the movable coil.

This deflecting torque is opposed by the spring restoring torque, T_S, given by

$$T_S = c\theta \tag{6.2}$$

where c is the spring stiffness and θ is the angular deflection. Neglecting friction, the resultant unbalanced torque on the coil is $T_D - T_S$. This is equal to the product of moment of inertia, J, and angular acceleration $d^2\theta/dt^2$, i.e.

$$BANI - c\theta = \frac{Jd^2\theta}{dt^2} \tag{6.3}$$

Considering connection of the recorder (coil) with the source as shown in Figure 6.6, the current I is given by

$$I = \frac{E_{TH} - E_b}{R_{TH} + R_R} \tag{6.4}$$

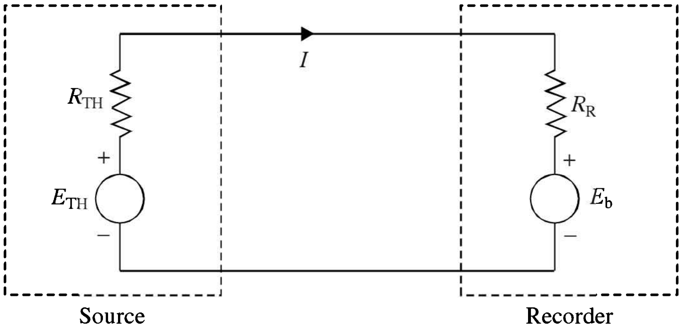


Figure 6.6 Connection of the recorder (coil) with the source.

where E_{TH} is the Thevenin voltage of the source, R_{TH} is the Thevenin resistance of the source, E_b is the back emf induced in the coil due to its motion in the magnetic field and R_R is the recorder (coil) resistance. E_b is given by

$$E_b = NBA\left(\frac{d\theta}{dt}\right) \tag{6.5}$$

Substituting Eqs. (6.4) and (6.5) into Eq. (6.3), we get

$$\frac{J}{c}\frac{d^2\theta}{dt^2} + \frac{(\mathrm{NAB})^2}{c(R_{TH} + R_R)}\frac{d\theta}{dt} + \theta = \frac{\mathrm{NAB}}{c(R_{TH} + R_R)}E_{TH} \tag{6.6}$$

If $\Delta\theta$ and ΔE_{TH} are deviations from the initial steady conditions, the transfer function of the recorder can be written from Eq. (6.6) in the standard format as

$$\frac{\Delta\theta}{\Delta E_{TH}}(s) = \frac{K}{(1/\omega_n^2)s^2 + (2D/\omega_n)s + 1} \tag{6.7}$$

where s is the complex frequency. K, D, and ω_n are given by

$$\text{Steady state sensitivity,} \quad K = \frac{\mathrm{NAB}}{c(R_{TH} + R_R)} \text{ rad/V} \tag{6.8}$$

$$\text{Damping factor,} \quad D = \frac{(\mathrm{NAB})^2}{2\sqrt{(cJ)}(R_{TH} + R_R)} \tag{6.9}$$

$$\text{Natural frequency,} \quad \omega_n = \sqrt{c/J} \text{ rad/s} \tag{6.10}$$

The typical frequency response of a recorder for different damping factors, D, is shown in Figure 6.7. This shows that best bandwidth is obtained at $D = 0.7$ and is therefore the target when an instrument is designed. The damping factor not only depends on the coil and spring constants but also on the total circuit resistance ($R_{TH} + R_R$). Adding a series or parallel resistance between the source and the recorder respectively, decreases or increases the damping factor. However, any reduction in D will be at the cost of sensitivity K [(Eq. 6.8)].

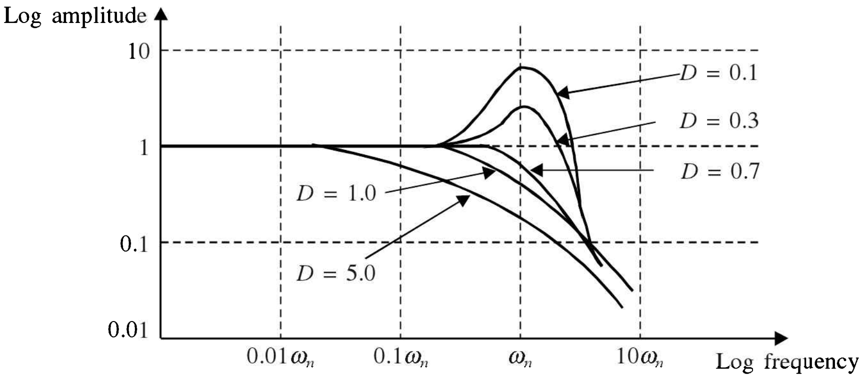


Figure 6.7 Frequency response of galvanometric recorder.

Reducing c gives a larger K but a lower ω_n; similarly increasing N and A also increases K but gives a larger moment of inertia J and again reduced ω_n. There is a limit as to how far the dynamic performance of a recorder can be improved. Other instruments such as ultra violet recorders can be used for high frequency signals.

6.2.2 Potentiometric/Servo/Null Balancing Recorders

Figure 6.8 shows a basic dc potentiometer servo recorder mechanism. The pen is attached to a wire (string) which is wound around two idler pulleys and a drive pulley on the shaft of a dc servomotor. The pen assembly is also linked with the wiper arm of potentiometer R_1 in such a way that the position of the wiper arm is proportional to the pen position. This linkage can be mechanical or ultrasonic.

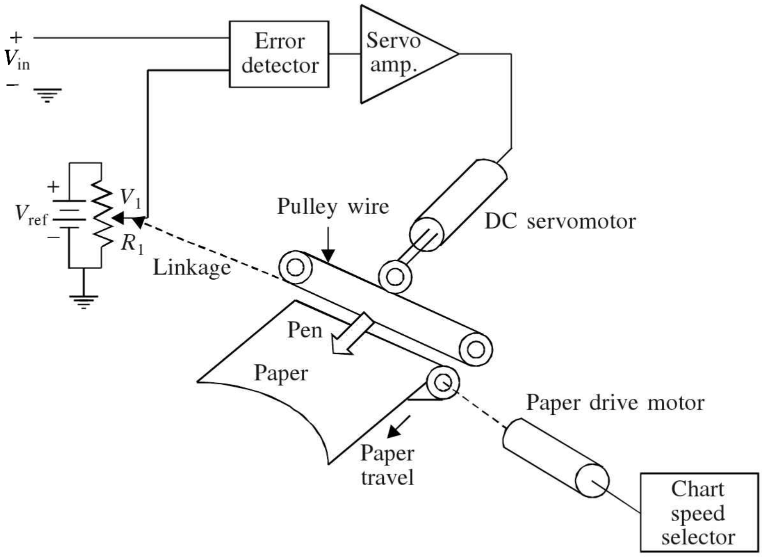


Figure 6.8 Servo recorder mechanism.

Voltage V_l is proportional to the position of the pen. When pen is at left-hand side of the paper $V_l = 0$, and when it is on right extreme $V_l = V_{ref}$. The error detector detects the difference between V_{in} and V_l and produces the output proportional to this difference. This is then amplified by the servo amplifier which drives the dc servo motor accordingly, thereby, controlling the movement of the pen. When $V_{in} = V_l$, error detector output becomes zero and hence pen movement is zero.

A paper drive motor forms a time base because it pulls the paper underneath the pen at a fixed, constant rate. The speed of this motor can be set by the chart speed selector. Some recorders use stepper motor in place of servomotor.

Ultrasonic pen position sensing

Pen position here is sensed from the propagation time of an ultrasonic pulse imparted to a magnetostrictive strip by a piezoelectric transducer as shown in Figure 6.9. This totally eliminates the sliding mechanical contacts and slide wire wear that limit the life of the potentiometers. Ultrasound velocity variation with temperature can also be cancelled by an offset calculation for highly accurate positioning.

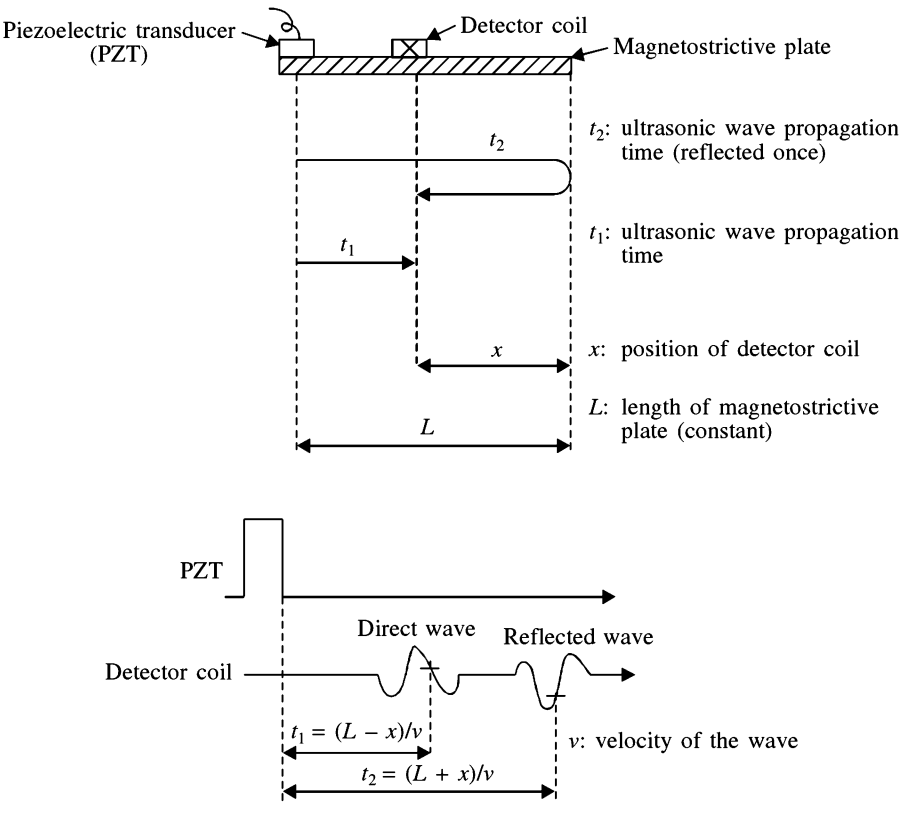

Piezoelectric transducer (PZT)
Detector coil
Magnetostrictive plate
t_2
t_1
x
L
t_2: ultrasonic wave propagation time (reflected once)
t_1: ultrasonic wave propagation time
x: position of detector coil
L: length of magnetostrictive plate (constant)
PZT
Direct wave
Reflected wave
Detector coil
$t_1 = (L - x)/v$
$t_2 = (L + x)/v$
v: velocity of the wave

Figure 6.9 Ultrasonic pen position transducer.

Magnetostrictive material is a ferromagnetic material (nickel or carbon steel) through which ultrasonic or elastic waves can transmit. In a magnetostrictive effect, magnetic field produces a small change in the physical dimensions of the material. Reverse of this is known as inverse magnetostrictive effect.

In this position sensing mechanism, a pulse applied to PZT produces change of magnetic field in the magnetostrictive material which generates ultrasonic (elastic) waves. The wave is detected by the detector coil which generates current or voltage pulse.

Typical strip chart recorder plot

Figure 6.10 shows a typical plot on the chart. Note that the vertical coordinate (time) is the X-axis of a graph and the horizontal coordinate (measured variable) is the Y-axis. The heavy vertical line in the centre of the chart is the curve plotted by the recorder (often called the chart record or trace).

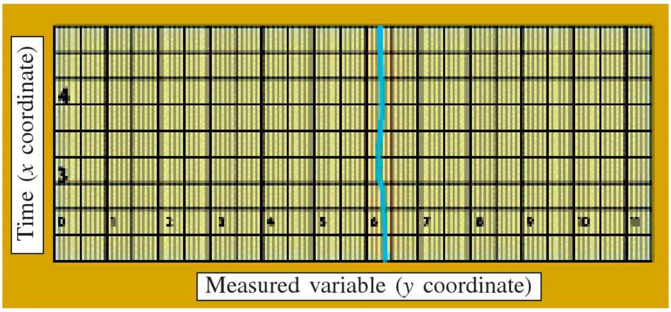


Figure 6.10 Typical strip chart recorder plot.

Chart speed

This is the rate at which recording paper in a strip chart recorder moves. It is expressed in inches/second or mm/second. These are switch selectable. If the chart speed is known, the frequency of the recorded signal can be calculated by

$$\text{Frequency} = \frac{\text{chart speed}}{\text{time base}} \tag{6.11}$$

Example 6.1

The chart speed of a recording instrument is 30 mm/s. One cycle of the signal being recorded extends over 5 mm (time base). Calculate the frequency of the signal.

Solution

$$\text{Frequency of the signal} = \frac{30 \text{ mm/s}}{5 \text{ mm/cycle}}$$

$$= 6 \text{ cycles/s}$$

Types of charts and accessories

Figure 6.11 shows a picture of various charts and marking pens used in the recorders.

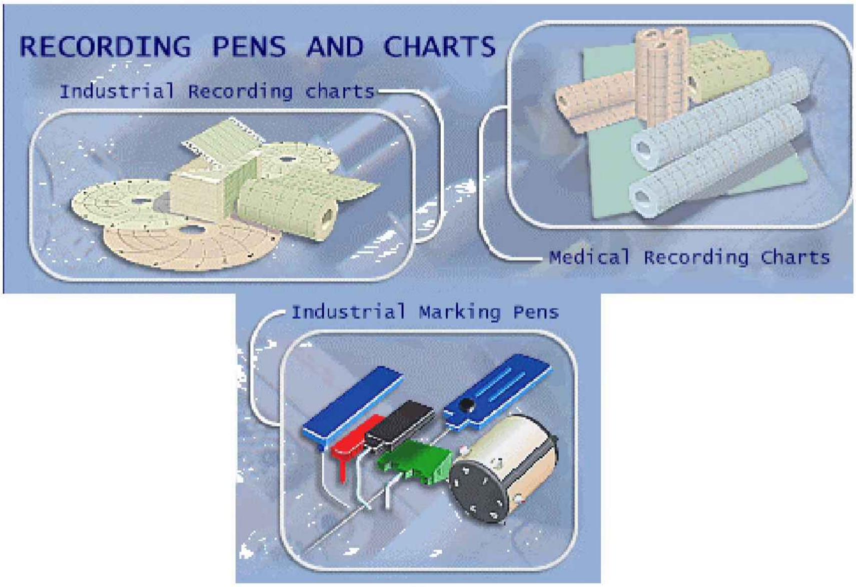


Figure 6.11 Various types of recording pens and charts.

The figure shows Z fold, roller type and circular charts. Normally these charts have markings indicating time and value of the parameter. For example, a temperature recorder chart would have chosen temperature markings and time. A circular chart has concentric circles and equally spaced arcs extended from centre to the rim of the chart. These are called time arcs. The time markings are placed along the circumference of the rim. The value of the measured parameter lies along the arcs. The concentric circles may be equally spaced or be at different distances from each other. The scale depends upon the linkage geometry of the instrument. Figure 6.12 shows a typical circular chart recorder trace.

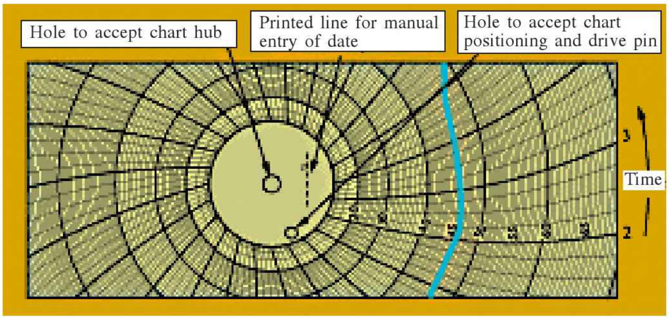

Hole to accept chart hub
Printed line for manual entry of date
Hole to accept chart positioning and drive pin
3
Time
2

Figure 6.12 Typical circular chart recorder trace.

Circular-chart coordinates were developed early in recorder history to utilize the convenience of a spring-wound clock and synchronous motor movements. Depending on application needs, the circular chart speed can be one rotation in 24 hours, one rotation in a week, or another arbitrary time interval. Note the advantage of a circular chart in giving at a glance a complete history of one or several variables over a period of time—albeit not with high-accuracy readings. Chart sizes are referred to by total diameter, such as 10″ or 12″, although full-scale pen movement is often much less. Because many plant operators are used to reading circular chart recorders, these are still readily available.

Multipoint recorders

The instruments which record changes of only one input parameter are called single-point recorders. Recorders are available that record many inputs simultaneously. For example, a multipoint recorder can record temperature, pressure, flow, level, etc. simultaneously. This feature can be performed either by having a recorder with multiple pens, each dedicated to an input (multipen recorder), or by using an auto selector switch and recording the selected input with its identification (multipoint recorder). Multipoint recorders can also use pens of different colours.

6.3 *X-Y* RECORDER

An *X-Y* recorder provides a graphic record of the relationship between two variables. Figure 6.13 shows the block diagram of a basic *X-Y* recorder.

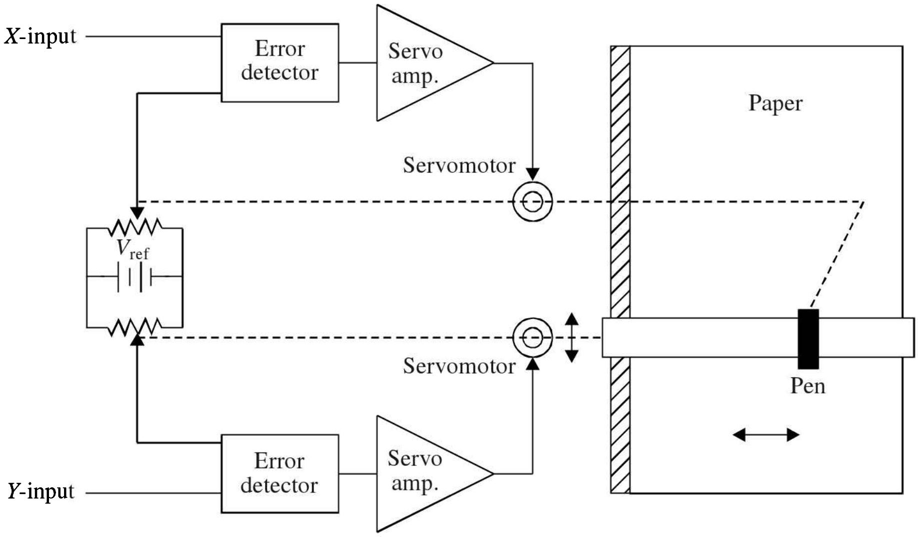


Figure 6.13 Block diagram of the basic *X-Y* recorder.

The *X*-input drives the pen in the *X* plane, while the *Y*-input drives the arm in the *Y*-plane. When *X*-input is present, the pen drive motor continues to drive the pen until error signal is zero. This is accomplished by the mechanical coupling between the pen drive motor and the movable arm of the potentiometer connected across the reference voltage. When the voltage at the movable arms equals the X-input voltage, the error signal equals zero and the pen drive motor stops turning. The same action takes place at the arm drive motion. The graphic record appears on a fixed paper chart.

XY recorders, though usually flatbed in design, are really a class by themselves. Most employ a potentiometric pen-positioning system, but some newer models feature a digital servo system. Input signals to the recorder can be analog or digital, with suitable conditioning as required. Common sizes for the chart paper are 8.5 by 11 and 11 by 17 inches.

Digital recorders use stepper motors to move the pens in X or Y directions. The stepper motor gives an incremental movement to the recording system for each small step of the data entered.

6.4 DATA LOGGERS

Based upon the way data is recorded and stored, there are two types of standalone data loggers. One type stores the data in memory to be retrieved at a later time; the other type automatically records the data on paper for immediate viewing and analysis. Many designs combine these two functions, with one or the other of primary importance.

There is a wide variety of software packages that can be loaded into PCs, digital control systems, or other computer-based devices to provide excellent data-logging and digital data storage capabilities. Also a large variety of standalone data loggers are available.

Note that hybrid recorders, as mentioned earlier, serve many of the functions formerly reserved for a separate data logger. Conversely, modern data loggers can also provide visual displays and digital read-outs. They can also perform complex mathematical operations on incoming data.

Figure 6.14 shows the block diagram of a data logger. Input signals (voltage, current, pulse etc.) are conditioned using proper signal conditioning circuits. These signals are multiplexed and are converted to their digital equivalents by the A-D converter. Microprocessor according to the program converts the data into engineering units, checks for the alarm limits, generates alarms, if required, stores the data, or outputs the data to the driver for digital display/printing or to D-A converter for analog recording on paper.

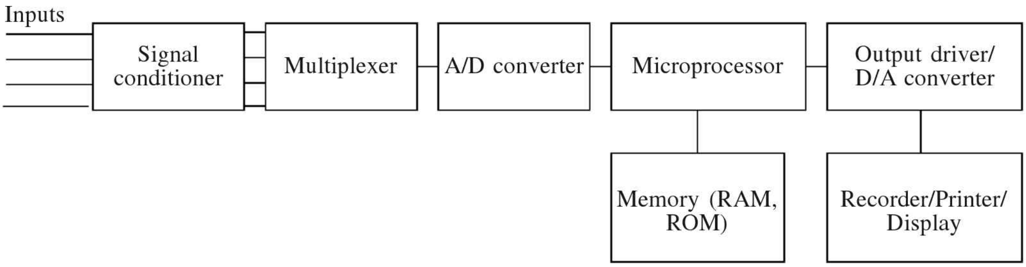


Figure 6.14 Block diagram of a data logger.

6.5 VIDEOGRAPHIC OR PAPERLESS RECORDERS

Videographic or paperless recorders can provide screen displays that look like a strip chart or a circular chart. Figure 6.15 shows two such paperless recorders.

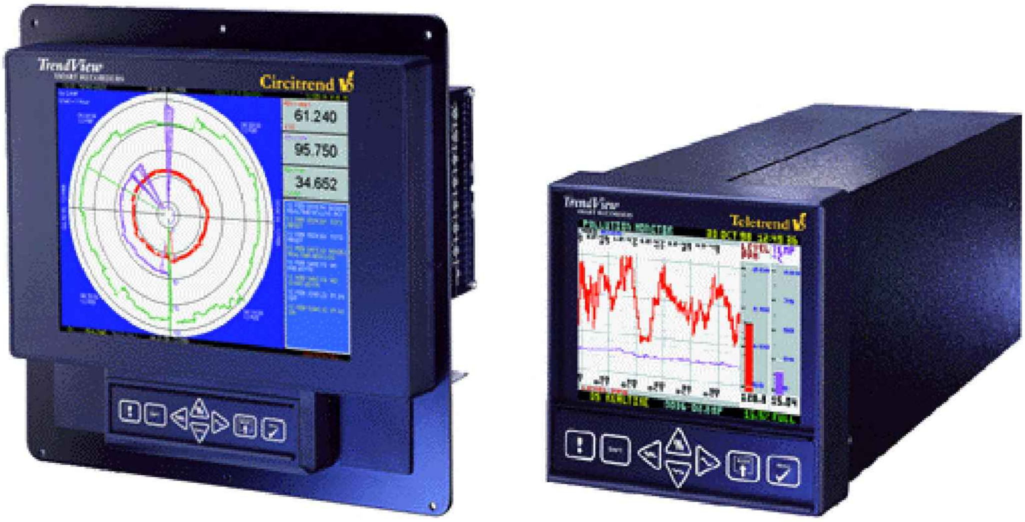


Figure 6.15 Paperless recorders with circular and strip chart displays.

Figure 6.16 shows paperless recorders displaying large information on the screen in various forms. These recorders are flexible in selecting displays and graphics, provide a variety of interfaces for the inputs, provide output ports for printing, standard current and voltage outputs, network facilities and easy programming.

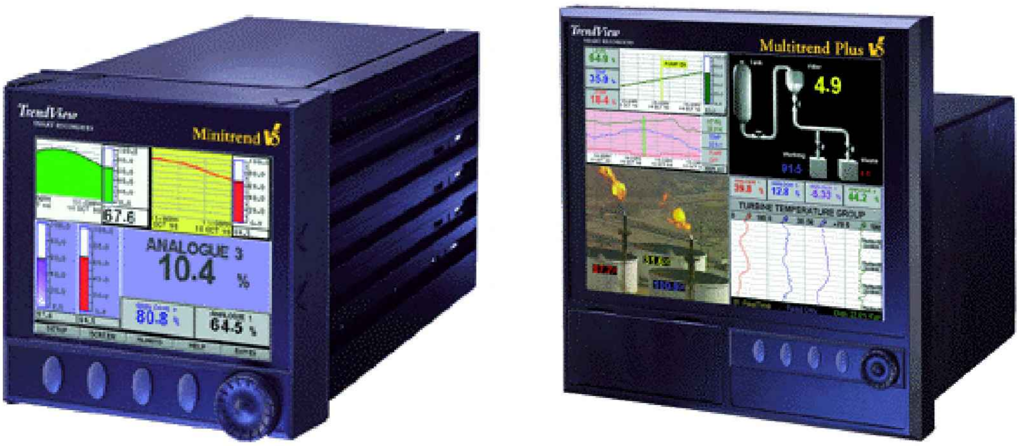


Figure 6.16 Paperless recorders with multi-type displays.

6.6 RECORDER SELECTION

Given the wide variety of instruments for recording or logging the acquired data, some key questions and basic points to consider in making a suitable selection are given below:

- How many inputs are to be recorded or logged? These are often called channels, points, or records.
- What types of inputs are involved? Voltages, T/C or RTD, 4-20 mA.
- Do different input types need to be handled by the same unit?
- What type of recording or logging is required—continuous single-point or multi-point? And if the latter, what minimum scan cycle is required?
- What degree of accuracy is required? Accuracy may be defined as the closeness to the actual signal that the measured value or trend position takes—stated in either a plus/minus percentage of full-scale or percentage of reading.
- Is a communication interface required to transmit measured data to a host computer?
- Is the instrument to be bench-top or to be mounted on a panel?
- What type of instrument power is available?
- Is log-type recording desired instead of or in addition to continuous trend recording?
- Is colour differentiation desirable for multiple trend lines?
- If the instrument is to perform alarm functions, how many alarm setpoints are needed per channel? What types of alarms are desired, e.g. threshold, rate, or delta? Are relays contacts required to handle external alarm outputs? If so, how many?
- What, if any, signal processing is required? These may include linear scaling (conversion to engineering units), thermocouple characterization, difference calculation, square root calculation (as for an orifice flowmeter).
- Are any higher-level math functions required? These may include absolute value, logarithm, exponential, maximum, minimum, time average, group average, summation, standard deviation, and integration functions.
- For conventional, paper-based recorders, what writing method is preferable? The mechanisms may be capillary ink, disposable felt-tip ink cartridge, dot printing, using an ink ribbon cassette or pressure sensitive paper, thermal moving head or stationary linear array, or rotating ink wheel.
- If a strip-chart recorder is used, is a Z-fold or roll type of chart desired? Z-fold has become the primary choice for process applications because of the ease with which an operator can review past traces without disrupting active recording.
- Should the chart speed be fixed or programmable?
- What kinds of annotations on the chart are desired? Tags, lists, engineering units, messages, scales, channel identification (numeric or alphanumeric), date and time, chart speed, and instantaneous *snapshot* values. Continuous writing instruments provide the annotation by a separate print head.

- For a strip-chart recorder, what chart width is desired—100 mm, 180 mm, or 250 mm, or any other?
- What secondary visual indications are desired? These may be analog bar graph in percentage of full scale, analog scale indication in percentage of full scale, digital channel (point) number and measured value, alarm status, engineering units.
- What operating modes will be required? Options include normal monitoring, which involves monitoring at a set scan interval and trend recording at a set chart speed, or logging at set intervals; print on alarm, which involves monitoring at a set scan interval but not trend recording or logging until an alarm condition occurs; and change on alarm, which allows tending or logging at a base chart speed or log interval when no alarm condition exists and an automatic switch to an alternate (usually faster) chart speed or log interval when an alarm condition exists.

QUESTIONS

6.1 Explain the operation of PMMC galvanometer type strip chart recorder.

6.2 What are the different writing mechanisms used in the PMMC type recorders? Compare their performance?

6.3 Describe the construction and operation of a two-channel UV recorder using single source and single magnet.

6.4 How is the dynamic response of a galvanometer recorder optimized? What are the limitations?

6.5 Explain the operation of a servo recorder using slide wire mechanical linkage. How is slide wire contact replaced in modern servo recorders?

6.6 Explain the basic mechanism used in dc potentiometer servo recorders. Discuss a mechanism to improve upon the sensing of pen position.

6.7 Explain the construction and function of basic *X*-*Y* recorders.

6.8 Draw the schematic of a data logger and explain its operation. How is it different from a recorder?

6.9 What are paperless or videographic recorders? What are the features available from these recorders?

6.10 List suitable points for the selection of a recorder.

6.11 Write typical specifications of a paperless recorder.

6.12 Write typical specifications of a strip chart recorder.

Chapter 7

Signal Generators

7.1 INTRODUCTION

A signal generator (source) is an important instrument of almost any instrumentation setup used in hardware design, debug, evaluation projects, etc. Second only to DMM, signal sources are perhaps the most universal class of electronic test instruments.

A signal generator's job is simply not to provide an ideal waveform. Often the instrument must add known, repeatable amounts and types of distortion to the signal it delivers. This characteristic is one of the signal source's strongest virtues, since it is often impossible to create predictable distortion exactly when and where it is needed using only the circuit itself. The response of the unit-under-test (UUT) in the presence of the distorted signals reveals its ability to handle stresses that fall outside the normal performance envelope.

A stimulus signal may take the form of a low distortion sine wave, a stream of logic pulses, a high frequency radio carrier wave, a mobile telephone transmission and many other formats. Traditionally, the task of producing these diverse waveforms has been filled by separate, dedicated signal sources, from ultra-pure audio sine-wave generators to multi GHz RF signal generators. Many a times the user has to custom design the source to meet the project requirements.

Commonly used circuits in the laboratory signal generators for audio frequency had been *RC* (phase shift and Wien bridge) oscillators and for radio frequency *LC* (Hartley, Colpitt and Crystal) oscillators. Modern instruments like arbitrary waveform generators, arbitrary function generators, data generators, etc. use other frequency synthesis techniques. We will study the oscillators and frequency synthesis techniques used in these instruments.

7.2 SINE-WAVE GENERATORS

Many systems require an input in the form of a periodic, usually sinusoidal, waveform. These generators use feedback oscillators. All feedback oscillators require some device or mechanism which provides gain combined with a feedback arrangement which sends some of the system's output back to be re-amplified after a suitable time delay. Figure 7.1 shows a general arrangement of the feedback arrangement.

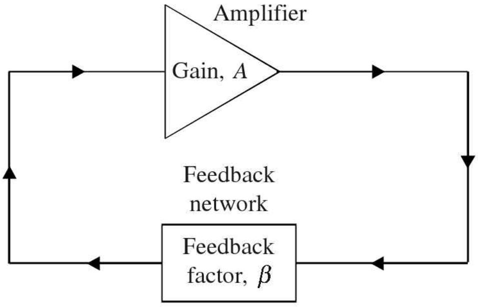


Figure 7.1 Feedback arrangement.

It shows an amplifier which has a voltage gain, A, whose output and input are connected via a feedback network. This network returns a fraction, β, of the output voltage to the amplifier's input. Both the amplifier's gain and the feedback factor are frequency dependent. In general, both the amplifier and the feedback network will alter the magnitude and the phase of the signal. To take this into account it is normal to treat both A and β as complex quantities.

The amplifier and feedback network form a loop. An initial signal fluctuation

$$V_{\text{in}}(t) = V_{\text{max}} \exp(-j\omega t)$$

at the amplifier's input will produce an output

$$V_{\text{out}}(t) = AV_{\text{max}} \exp(-j\omega t)$$

which, in turn, produces a new 'echoed' input signal

$$V'_{\text{in}}(t) = A\beta V_{\text{max}} \exp(-j\omega t)$$

back at the amplifier's input. ($A\beta$ is called the *loop gain*). This new input will, in turn, be amplified to produce a new echo at the input, and so on.

After n 'trips around the loop' the amplitude of the newest echo will be

$$|V_{\text{in}}| = |A\beta|^n \, |V_{\text{max}}|$$

By looking at this expression we can see that if $|A\beta| < 1$ the echoes will fade away. However, if we arrange $|A\beta| \geq 1$, then the size of the echoes tends to grow with time (or at least remain constant if we arrange that $|A\beta| = 1$) and will be limited by the value of the power supply. This is called *regenerative* or *positive* feedback. The gain of the amplifier with feedback, $A_f = A/(1 - A\beta)$ tends to infinity if $A\beta = 1$.

As a result, we find that an initial signal produces a sustained, repeating signal whose amplitude does not fade away with time provided that we can ensure that loop gain magnitude

$$|A\beta| \geq 1$$

and the phase shift

$$\angle A\beta = \angle A + \angle \beta = 2\pi n \qquad \text{where } n = 1, 2, 3, \ldots$$

That is, the phase shift around the loop must be an integer number of cycles.

Each delayed echo or *cycle* of fluctuation will append itself onto the tail of the previous fluctuation with the same sinusoidal phase. As a result, provided that these two expressions are satisfied, we only have to 'give the system a kick' by providing the initial cycle of input.

These expressions for loop gain magnitude and phase shift taken together are called the *Barkhausen criterion*. Any system which satisfies this criterion is able to oscillate at any frequency for which the expressions are both true. Note that, in practice, phase angle expression is usually only satisfied for a one or more discrete frequency values, hence the system can only oscillate at these specific frequencies. It should also be noted that a system satisfying the criterion does not guarantee that it will actually oscillate. The process has to be started by an initial small fluctuation of the correct frequency. If this starting kick is absent, the system may just sit in a quiescent state. Fortunately, any small fluctuation which contains some power at the frequency, f, will start a sequence of steady or growing oscillations at that frequency. In practice, this means that we do not usually have to provide a specific input to start the process. The pulse generated by switching on the oscillator's amplifier (the gain source) is usually enough to get things going. If not, the random noise present in all real physical systems can often provide the required starting kick.

In general, as an oscillator tends to start up, its oscillation amplitude grows (usually rapidly), until limited by some process or feature of the system. The action of this limiting process is to reduce the effective loop gain until its modulus is unity. The oscillation then continues with a waveform of an essentially constant amplitude. This is the example of a simple harmonic oscillator; hence it generates a sine wave. Other forms of oscillators can produce other types of waveforms—square waves, triangle waves, etc.

Frequency stability

The frequency stability of an oscillator is a measure of the degree to which a constant frequency output is approached. The better the frequency stability is, the closer the output would be to a constant frequency.

Frequency instability (variations above and below the desired output frequency) may be caused by device characteristics or by variations in the external circuit elements. When the operation of the circuit falls into the nonlinear portion of the characteristic curve, the device's parameters (voltages and currents) vary. These parameters are basic to the stable frequency of the oscillator. Operating frequency variations may occur with changes in the bias voltages. Thus, a constant supply voltage is a prime requirement for good frequency stability.

Further, the loop gain phase shift varies with frequency since both A and β are frequency dependent. The rate of change of phase with frequency, $d\phi/d\omega$, should be as large as possible for the frequency stability. This is shown in Figure 7.2. Let ω_0 be the frequency

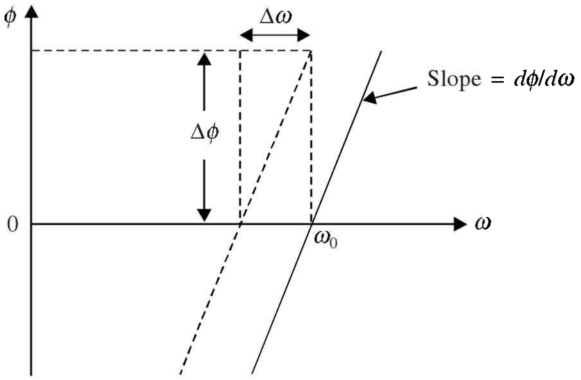


Figure 7.2 Loop gain phase shift variation.

when the loop gain phase shift, ϕ is zero. If due to change in some component value, phase changes by $\Delta\phi$, then the change in frequency $\Delta\omega_0$ is minimum if $d\phi/d\omega$ is large.

Amplitude stability

The amplitude stability of an oscillator indicates the amount by which the actual output amplitude varies from the desired output amplitude. The amplitude of the waveform changes with the gain of the amplifier. If due to change in component values, the gain changes, the amplitude will also change. Various gain control techniques are used to keep the amplitude constant.

Variations in the supply voltage may also cause changes in amplitude. Regulation of the supply voltage ensures good amplitude stability.

7.2.1 Audio Frequency Signal Generators

Audio frequency signal generators provide, normally, sinusoidal waveforms within the frequency range of 20 Hz to 20 kHz. Their output impedance is selected between 50 and 600 ohms depending on the load. The block diagram of a simple signal generator is shown in Figure 7.3.

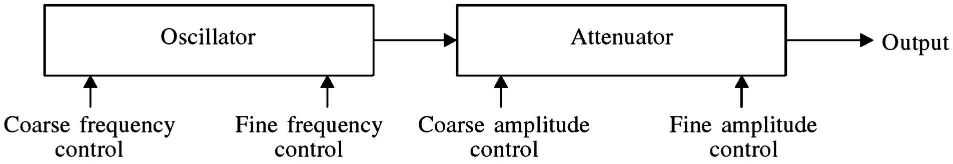


Figure 7.3 Block diagram of a signal generator.

Resistance-capacitance (*RC*) feedback oscillators are used to generate audio frequency signals. Most commonly used oscillators are Wein bridge and *RC* phase shift oscillators.

Wein bridge oscillator

Figure 7.4 shows the circuit diagram of a basic Wein bridge oscillator using an opamp.

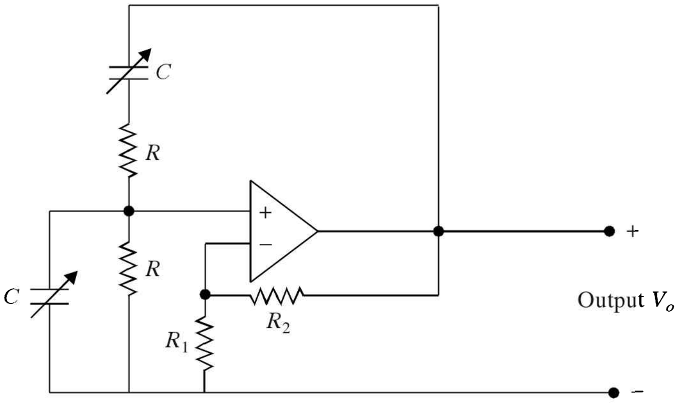


Figure 7.4 Wein bridge oscillator.

Since the opamp is being used as a noninverting amplifier, the gain of the amplifier is positive and is equal to

$$A = 1 + \frac{R_2}{R_1}$$

It means that the output and input of this amplifier are in phase ($\angle A = 0$). A series combination of resistance R and capacitance C and a parallel combination of R and C form the feedback network, which feeds back a portion of the output voltage V_o. The amount of feedback (feedback factor β) is given by

$$\beta = \frac{1}{3 + j\left(\omega RC - \dfrac{1}{\omega RC}\right)}$$

For oscillations $\angle A + \angle\beta = 0$. Since $\angle A = 0$, $\angle\beta$ should also be 0. Therefore, β should be real, which is possible if

$$\omega = \frac{1}{RC}$$

This gives the frequency of oscillations, $f_0 = \frac{1}{2\pi}RC$.

For $\omega = 1/RC$, β should be 1/3. Now for $A\beta = 1$,

$$A = 1 + \frac{R_2}{R_1} = 3$$

which implies, $R_2 = 2R_1$.

The resistances R_1 and R_2 must be carefully selected to provide a gain of exactly 3. Since this is not possible in real life (due to component tolerances and other problems), some form of amplitude stabilization is needed to ensure that the gain is automatically corrected. There are a number of possibilities as outlined below:

- **Thermistor.** A negative temperature coefficient thermistor is used in place of R_1. This automatically maintains the ratio of R_2 by R_1 as the temperature varies.
- **Low power lamp.** The filament of the lamp has a positive temperature coefficient, but requires more power than the thermistor. This is used in place of R_2.
- **LDR.** A light dependent resistor has a very high voltage limit before getting distorted, and can thus produce very good results. Although it requires more additional circuitry than the thermistor or lamp, the result is worth the effort. The LDR can be used either in place of R_1 or R_2, but it is more convenient to use it for R_2 because the circuit is then simpler, and the voltage across the device is minimized.
- **FET.** A field effect transistor works quite well as a voltage controlled resistor, but has a limited peak voltage, so the level must be kept low if distortion is to remain within respectable limits.

The Wien bridge oscillator has low distortion, excellent frequency stability, ease of change of frequency, wide frequency range (5 Hz to 500 kHz). The upper limit is fixed by the amplitude and phase shift characteristics of the amplifier. The lower limit is fixed by the practical limits on the size of range selecting resistors.

RC phase shift oscillator

Figure 7.5 shows the circuit diagram of a basic *RC* phase shift oscillator using an opamp. Since the opamp is being used as an inverting amplifier, the gain of the amplifier is negative and is equal to

$$A = -\frac{R_2}{R_1}$$

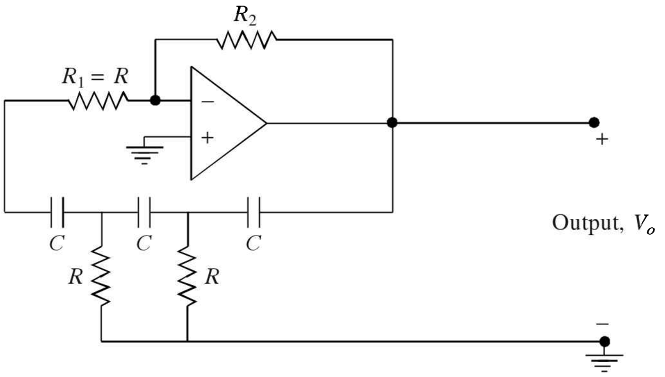


Figure 7.5 *RC* phase shift oscillator.

It means that the output and input of this amplifier are out of phase ($\angle A = 180°$). Three combinations of R and C form the feedback network, which feeds back a portion of the output voltage V_o. The amount of feedback (feedback factor β) is given by

$$\beta = \frac{1}{1 - 5\alpha^2 - j(6\alpha - \alpha^3)}$$

where $\alpha = 1/\omega RC$.

For oscillations, $\angle A + \angle\beta = 0$. Since $\angle A = 180°$, $\angle\beta$ should also be 180°. (Each RC combination provides a phase shift of 60°). Therefore, β should be real, which is possible if

$$6\alpha - \alpha^3 = 0, \qquad \text{or } \omega = \frac{1}{\sqrt{6}RC}$$

This gives the frequency of oscillations, $f_0 = \dfrac{1}{2\pi\sqrt{6}RC}$

For $\omega = 1/(\sqrt{6}\,RC)$, $\beta = -1/29$. Now for $A\beta = 1$,

$$A = -\frac{R_2}{R_1} = -29$$

which implies, $R_2 = 29R_1$.

R_1 and R_2 must be carefully selected to provide a gain of 29. Generally, R_2 is made variable and is adjusted to get sustained oscillations. Amplitude stabilization can be achieved by the techniques used in Wein bridge oscillator.

For change of frequency, either R or C or both, can be made variable. Note that all three resistors (R) or capacitors (C) should be ganged so that the same change in R or C is achieved.

Using more than three RC sections reduces the overall signal loss within the network. This is because additional RC sections reduce the phase shift necessary for each section, and the loss for each section is lowered as the phase shift is reduced. In addition, an oscillator that uses four or more RC networks has more stability than the one that uses three RC networks. In a four-part RC network, each part shifts the phase of the feedback signal by approximately 45° to give the total required 180° phase shift. The frequency of oscillation is given by

$$f_0 = \frac{1}{2\pi\sqrt{2n}RC}$$

However, the cost increases and the problem of variation of all resistors or capacitors becomes prominent.

The RC phase shift oscillator has a simpler circuit than that of the Wein bridge, but gives high distortion. Frequency stability is poorer than that of the Wein bridge. Change of frequency is inconvenient. This oscillator is therefore useful only for non-critical applications, particularly at low and medium frequencies.

7.2.2 Radio Frequency Signal Generators

Radio frequency signal generators are used to provide sinusoidal waveforms above 20 kHz. Their output impedance is selected between 50 and 600 ohms depending on the load.

Instead of *RC* feedback networks, these generators use inductance-capacitance (*LC*) tuned circuits as the feedback networks. Most commonly used oscillators are Hartley, Colpitts and Crystal oscillators. Figure 7.6(a) shows the general configuration of these oscillators, and Figure 7.6(b) gives the equivalent circuit of Figure 7.6(a). In this circuit, A_V is the voltage gain of the amplifier, and R_o is the output resistance of the amplifier. Input resistance of the amplifier is assumed to be infinite. Z_1, Z_2 and Z_3 are the impedances of the components connected in the respective branches of the feedback network.

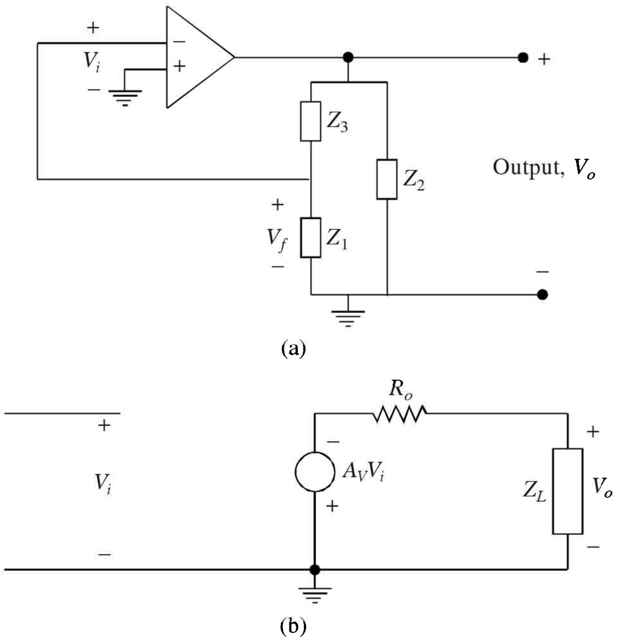


Figure 7.6 (a) General configuration of *LC* oscillators and (b) its equivalent circuit.

Without any feedback, the gain of the amplifier, with load Z_L, is given by

$$A = -\frac{A_V Z_L}{R_o + Z_L} \qquad \text{where } Z_L = Z_2 \parallel (Z_1 + Z_3)$$

The feedback factor, β, is obtained by finding the feedback voltage, V_f, in terms of V_o, i.e.

$$\beta = \frac{Z_1}{Z_1 + Z_3}$$

The loop gain (calculated by multiplying A and β) should be equal to 1, i.e.

$$A\beta = -\frac{A_V Z_1 Z_2}{R_o(Z_1 + Z_2 + Z_3) + Z_2(Z_1 + Z_3)} = 1$$

Since the feedback elements are either inductors or capacitors, impedance Z is equal to $j\omega L$ for the inductor and equal to $-j/\omega C$ for the capacitor. Therefore, the first term in the denominator is imaginary and the second is real. For $A\beta = 1$, the first term in the denominator should be zero, i.e.

$$Z_1 + Z_2 + Z_3 = 0 \quad \text{or} \quad Z_1 + Z_3 = -Z_2$$

This equation gives the frequency of oscillations. The value of feedback factor β, is then equal to

$$\beta = -\frac{Z_1}{Z_2}$$

From the loop gain equation, we get

$$A_V = \frac{Z_2}{Z_1} \text{ and } A = -\frac{Z_2}{Z_1}$$

To satisfy these equations, Z_1 and Z_2 should be of same type, that is either inductive or capacitive, and Z_3 should be of opposite type, that is if Z_1 and Z_2 are inductive, Z_3 should be capacitive or vice versa.

It may be noted that the phase shift equation for oscillations is satisfied since the amplifier gives 180° phase shift and other 180° is provided by the β network, thus, making the total phase shift equal to 2π.

In the Hartley oscillator, Z_1 and Z_2 are inductive and Z_3 capacitive, whereas in the Colpitts oscillator, Z_1 and Z_2 are capacitive and Z_3 inductive.

Hartley oscillator

Figure 7.7 shows the circuit diagram of the Hartley oscillator.

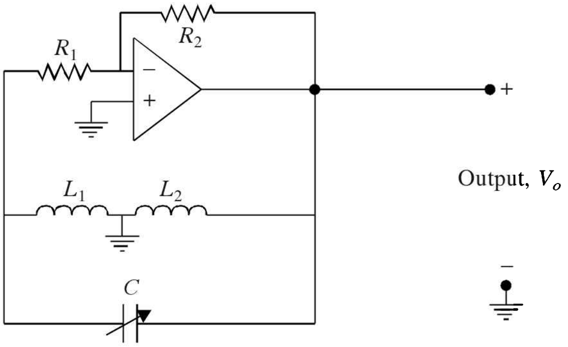


Figure 7.7 Hartley oscillator.

Here $Z_1 = j\omega L_1$, $Z_2 = j\omega L_2$, and $Z_3 = -j/\omega C$, where L_1 and L_2 are parts of a centre-tapped inductor. The frequency of oscillation, the gain of the amplifier and the feedback factor, can then be calculated from the above equations and are given as

$$f_0 = \frac{1}{2\pi\sqrt{(L_1 + L_2)C}}; \quad A = -\frac{L_2}{L_1}; \quad \beta = -\frac{L_1}{L_2}$$

The resistors R_2 and R_1 should be adjusted for the gain of A_V.

Note that inductors and capacitor form a tuned circuit, which provides oscillations at frequency, f_0.

In the Hartley oscillator the frequency is simply varied by the value of C, the output amplitude remains constant when tuned over the frequency range, and the feedback ratio of L_1 to L_2 remains constant. The output is rich in harmonic content and therefore not suitable where a pure sine wave is required. The opamp used should have a high slew rate (V/μs) and high bandwidth, otherwise BJT or FET should be used in lieu.

Colpitts oscillator

Figure 7.8 shows the circuit diagram of the Colpitts oscillator.

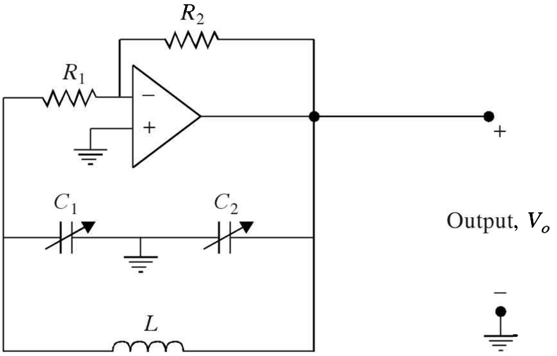


Figure 7.8 Colpitts oscillator.

Here $Z_1 = -j/\omega C_1$, $Z_2 = -j/\omega C_2$, and $Z_3 = j\omega L$, where C_1 and C_2 are ganged to maintain the same ratio. The frequency of oscillation, the gain of the amplifier and the feedback factor can then be calculated from the above equations as

$$f_0 = \frac{1}{2\pi\sqrt{\frac{C_1 C_2}{C_1 + C_2} L}}; \quad A = -\frac{C_1}{C_2}; \quad \beta = -\frac{C_2}{C_1}$$

The resistors R_2 and R_1 should be adjusted for the gain of A_V. Here also the inductor and capacitors form a tuned circuit which provides oscillations at frequency f_0.

In the Colpitts oscillator the frequency is varied by ganged capacitors C_1 and C_2. No centre-tapped inductor is required. The output amplitude remains constant when tuned over the frequency range. Also, the feedback ratio of C_2 to C_1 remains constant. The oscillator has comparatively good waveform purity since C_1 and C_2 provide low impedance path for harmonics. The opamp should have high slew rate (V/μs) and high bandwidth, otherwise BJT or FET should be used in lieu.

Crystal oscillator

Crystal oscillators are oscillators where the primary frequency determining element is a quartz crystal. Because of the inherent characteristics of the quartz crystal, the crystal oscillator can be held to extreme accuracy of frequency stability. Temperature compensation may be applied to crystal oscillators to improve thermal stability of the crystal oscillator.

Crystal oscillators are usually fixed frequency oscillators, where the stability and accuracy are the primary considerations. For example, it is almost impossible to design a stable and accurate *LC* oscillator for the upper HF and higher frequencies without resorting to some sort of crystal control.

Crystals exhibit a characteristic known as *piezoelectric effect*. The piezoelectric effect is the property of a crystal by which the mechanical forces produce electrical charges and, conversely, electrical charges produce mechanical forces. This effect is a form of oscillation similar to a tuned circuit. The piezoelectric effect can be seen in a number of crystal substances. The most important of these are the minerals quartz (silicon dioxide) and Rochelle salt. Although quartz does not exhibit the piezoelectric effect to the degree that Rochelle salt does, quartz is used for frequency control in oscillators because of its greater mechanical strength.

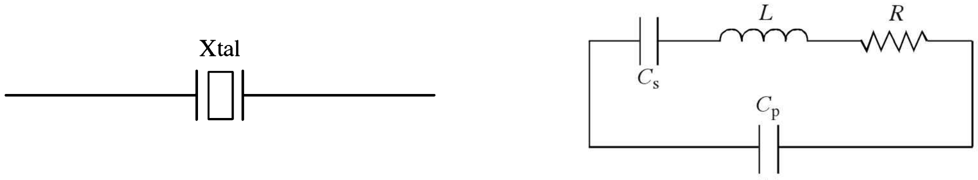
The crystals used in oscillator circuits are thin sheets, or wafers, cut from natural or synthetic quartz and ground to a specific thickness to obtain the desired resonant frequency. The crystals are mounted in holders which support them physically and are equipped with electrodes by which voltage is applied. The holder must allow the crystals, the freedom for vibration.

The frequency for which a crystal is ground is referred to as the *natural resonant frequency* of the crystal. If we apply an ac voltage to the quartz crystal, it will resonate and produce an electrical resonance at the natural resonant frequency that can be amplified. The natural resonant frequency is very stable. In addition, the resonance has very high Quality Factor, Q, ranging from 10,000 to as high as 2,000,000. The high Q causes the frequency stability to be much greater than that of an ordinary *LC* tuned circuit (which has Q of less than 200).

The resonant frequency is primarily determined by the physical dimensions of the crystal. However, cuts (the plane or angle of the slice through the main crystal) provide different frequency ranges and characteristics. By proper selection of the type of the cut, dimensions of the plate (plates are the electrical contacts to the crystal) and mode of vibration, it is possible to obtain crystals with resonant frequency from 5 kHz to 75 MHz. For higher frequencies, crystal plates become too thin and fragile and also more susceptible to frequency changes with temperature. But if the crystals are operated at odd harmonics (overtones) of the fundamental frequency, they can be used at higher frequencies (up to 200 MHz). We will assume the fundamental mode operation of the crystal.

Equivalent circuit of a crystal

Figure 7.9 shows the equivalent circuit of a quartz crystal near its fundamental resonance.



(a) (b)

Figure 7.9 (a) Quartz crystal symbol and (b) equivalent circuit of quartz crystal.

The components C_s, L, and R represent motional arm and mechanical behaviour of the crystal. C_p represents electrical behaviour (electrodes on the plate) of the crystal and the holder (which holds the crystal in the enclosure). Typical values of these components respectively are 0.018 pF, 22 mH, 30 Ω, and 4.5 pF.

The Q of the crystal is defined as

$$Q = \frac{X_L}{R} = \frac{1}{RX_{C_s}}$$

where X_L is reactance of the inductor and X_{C_s} is the reactance of the capacitor at the frequency of operation. For high Q, R is very small. For Q greater than 10, Q can also be defined as

$$Q = \frac{f}{\text{BW}}$$

where f is the resonant frequency and BW is 3 dB bandwidth.

Resonant frequencies of a crystal

Figure 7.10 shows the impedance frequency plot of the crystal. The crystal has two resonant frequencies characterized by zero phase shift. These are series resonant frequency, f_s, and parallel or anti-resonance frequency f_a. When a crystal is operating at series resonance, it looks purely resistive with minimum impedance (maximum current) and the reactances of the inductor and the capacitor are equal ($X_L = X_{C_s}$). The series resonance frequency is given by

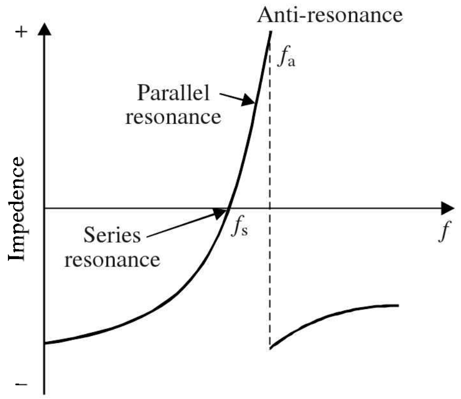


Figure 7.10 Impedance frequency plot of quartz crystal.

$$f_s = \frac{1}{2\pi\sqrt{LC_s}}$$

The reactance of the shunt capacitance, X_{C_p} is in parallel with R. At resonance, $X_{C_p} >> R$, thus the crystal appears resistive at the value of approximately equal to R.

When a crystal is operating at anti-resonance frequency, it again looks purely resistive with maximum impedance (minimum current), and the reactances of the inductor and the series combination of capacitors, C_s and C_p, are equal. The anti-resonance frequency is given by

$$f_a = \frac{1}{2\pi\sqrt{L\dfrac{C_sC_p}{C_s + C_p}}}$$

Note that $f_s < f_a$ and the specified crystal frequency, f_{XTAL}, satisfies the equations

$$f_s < f_{XTAL} < f_a$$

The impedance between f_s and f_a looks inductive and this area of frequency is called *parallel resonance mode*. The frequency of operation in this mode also depends on the load on the crystal. Generally, the crystal manufacturers specify the load capacitance C_L for parallel resonant crystals. C_L will appear in parallel with C_p.

Circuit: crystal oscillator

Since in parallel resonance mode between frequencies f_s and f_a, the crystal behaves as an inductor, we can simply replace the inductor in the Colpitts oscillator by the crystal to get a crystal oscillator. Figure 7.11 shows the circuit of a crystal oscillator.

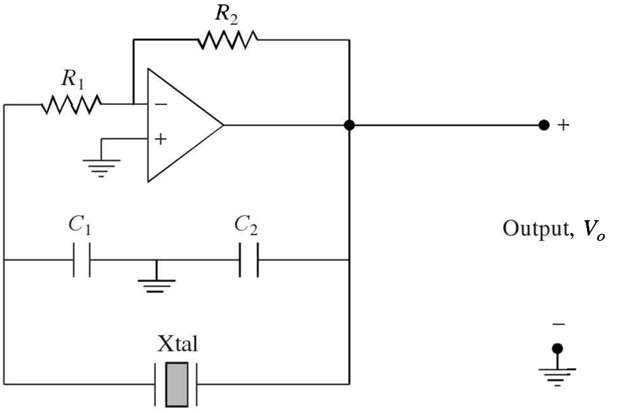


Figure 7.11 Crystal oscillator circuit.

The circuit oscillates at the resonant frequency of crystal inductance in series equivalent of C_s and $[C_p + (C_1C_2/C_1 + C_2)]$. Since C_s is very small compared with other capacitors, the frequency is

$$f_0 = f_s = \frac{1}{2\pi\sqrt{LC_s}}$$

A small trimmer capacitor may be placed in series with the crystal to enable small adjustments to the crystal frequency. This capacitor may typically lie between 20–30 pF and may vary the frequency by a ± few kHz to maintain predictable operation. Adjusting the crystal frequency this way is called *pulling* the crystal.

Temperature coefficient: crystal oscillator

The change in crystal frequency with temperature is called its *temperature coefficient.* A crystal might have positive, negative, or zero temperature coefficient. It is expressed in hertz per megahertz per degree Celsius, or in parts per million (ppm) with degree Celsius understood.

When crystal stability is important, the crystal is enclosed in a temperature controlled oven. Crystal is kept at constant temperature by a heating and feedback mechanism to keep the constant temperature. Ovens are mostly used in the environment where the changes in temperature are extreme.

Features: crystal oscillator

A crystal oscillator has high frequency stability and high Q. Since the ratio C_s/C_p is very small, the coupling between the crystal resonator and the external circuit is small. Also since the tuning range is limited, crystal oscillators can be used at radio frequencies only.

7.3 NON-SINUSOIDAL GENERATORS

Non-sinusoidal oscillators generate complex waveforms such as square, triangular, pulse, sawtooth, etc. Because the outputs of these oscillators are generally characterized by a sudden change or relaxation, these oscillators are often called *relaxation oscillators* or *multivibrators.*

Multivibrators work in either of the two output states, ON or OFF (High or Low). There are three types of multivibrators:

1. **Bistable multivibrator.** This multivibrator is also called *flip-flop* or *latch.* In this multivibrator both the states are stable. It remains in one of the states until a trigger is applied, it then flips to the other state and remains there until another trigger pulse is applied [see Figure 7.12(a)]. Three trigger pulses change the output three times. After the third pulse, the output remains high.

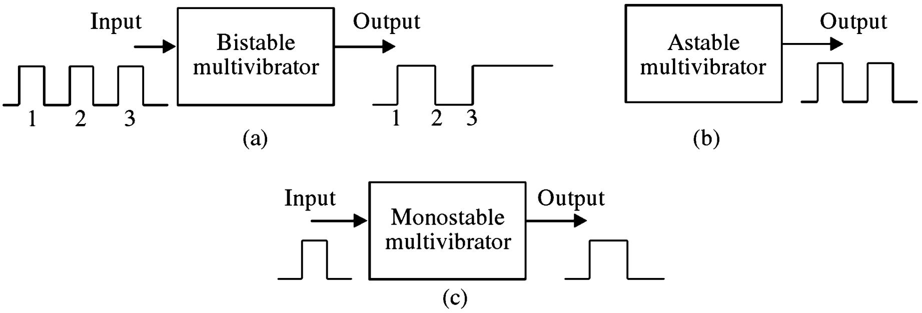


Figure 7.12 The operation of (a) bistable, (b) astable, and (c) monostable multivibrator.

2. **Astable multivibrator.** This multivibrator generates a free-running signal or a clock signal. None of the two states is stable [see Figure 7.12(b)]. The output continuously changes from high to low and low to high.
3. **Monostable multivibrator.** The output in this multivibrator is stable in one state and unstable in the other. A trigger causes the output to make a transition from the stable to the unstable state, where it remains for a length of time, and then reverts to the stable state. Monostable multivibrators are sometimes referred to as ***one-shots*** since one "shot" from the trigger creates a change in the output [see Figure 7.12(c)]. A short input trigger pulse produces a long output pulse which then returns to the original state.

The multivibrators are designed using discrete transistors, operational amplifiers, timers or digital ICs. We will now study the circuits of multivibrators employing operational amplifiers.

Bistable multivibrator

Figure 7.13 shows the basic circuit of a bistable multivibrator which is a comparator circuit where regenerative (positive) feedback is provided by the R_1 and R_2 feedback network with a feedback factor, $\beta = R_1/(R_1 + R_2)$. The output, v_o is either $+V_{sat}$ (+ saturation voltage) or $-V_{sat}$ (– saturation voltage) depending upon whether v_+ is greater than v_i or less than v_i. The value of v_+ is, therefore, either equal to $+\beta V_{sat}$ or $-\beta V_{sat}$. The transfer characteristics v_i versus v_o are shown in Figure 7.14.

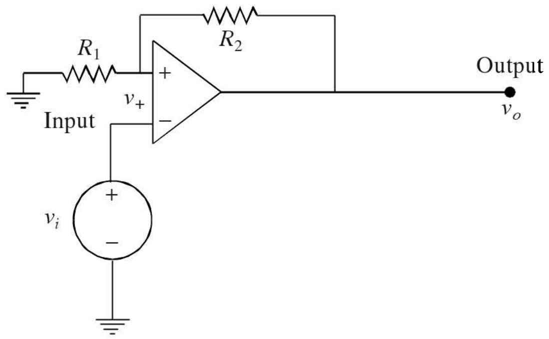


Figure 7.13 The circuit of a bistable multivibrator.

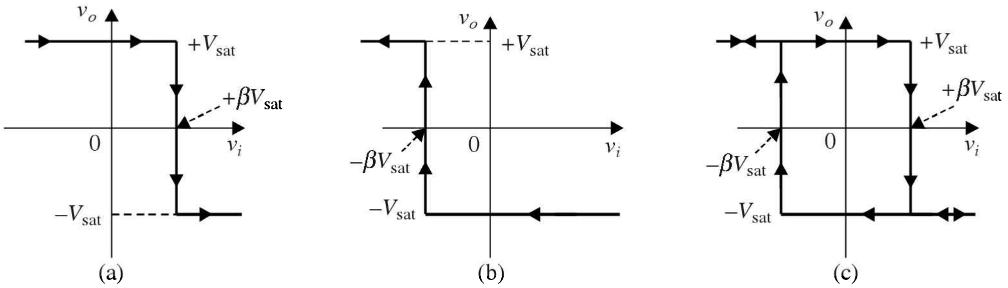


Figure 7.14 Transfer characteristics (a) increasing v_I, (b) decreasing v_I, and (c) complete.

It is observed from Figures 7.14(a) and (b) that the output voltage remains at the acquired state, $+V_{sat}$ or $-V_{sat}$ after the input voltage exceeds voltage equal to $-\beta V_{sat}$ or $+\beta V_{sat}$ respectively. These states are two stable states of the multivibrator. To change the state, we need to apply a pulse (trigger) to the input which should make $|v_I| \geq |\beta V_{sat}|$. The pulse can be removed once the regenerative process is competed and the output stays in that state.

As is observed from Figure 7.14(c), the circuit changes state at different values of the input depending upon whether input is increasing or decreasing. Such a behaviour is called *hysteresis*. The width of the hysteresis, V_H, is given by

$$\begin{aligned} V_H &= +\beta V_{sat} - (-\beta V_{sat}) \\ &= 2\beta V_{sat} \end{aligned}$$

where $\beta = R_1/(R_1 + R_2)$.

Thus a bistable multivibrator circuit is a comparator with hysteresis. This is also called the *Schmitt trigger circuit.*

Schmitt trigger circuit

This circuit is used to generate square or rectangular waveforms from a sinusoidal waveform and to condition any variable input. Let us apply a sinusoidal input to the Schmitt trigger circuit of Figure 7.13. The output of the circuit will be a square or rectangular waveform depending upon the value of βV_{sat}. Figure 7.15 shows the input and output waveforms. $+\beta V_{sat}$ value is called the *upper threshold level* V_{UT}, and $-\beta V_{sat}$ value is called *lower threshold level* V_{LT}.

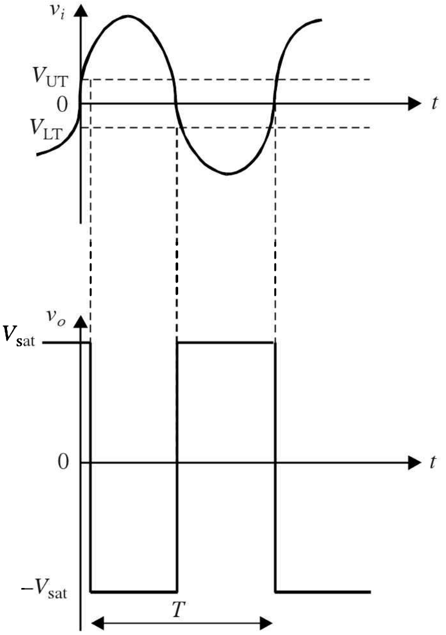
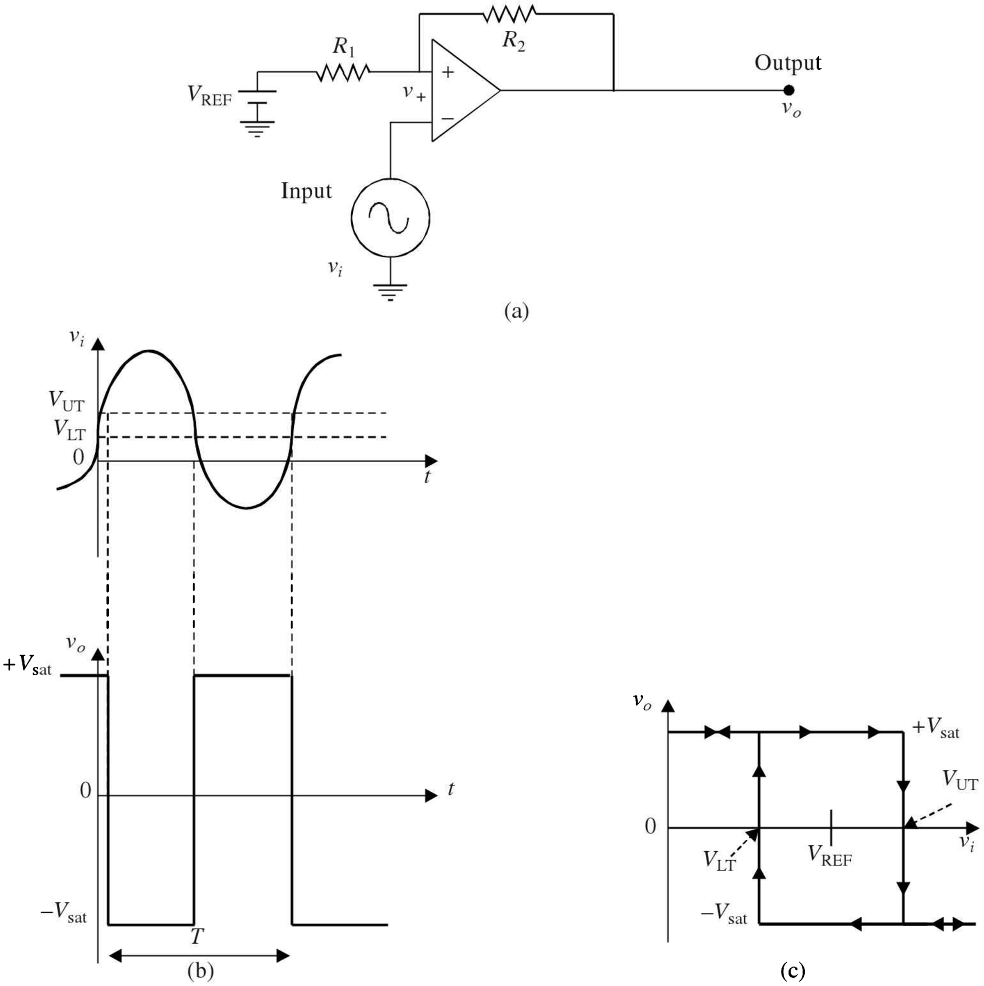


Figure 7.15 Input and output waveforms of Schmitt trigger.

It can be observed that the values of V_{UT} and V_{LT} are symmetrical about zero for same β and V_{sat}. We can shift the point of symmetry as well as change the values of V_{UT} and V_{LT} by connecting a dc voltage source between R_1 and ground. The modified circuit, the input and output waveforms, and the hysteresis are shown in Figure 7.16.



(a)

(b)

(c)

Figure 7.16 (a) Modified Schmitt trigger circuit, (b) input and output waveforms, and (c) hysteresis.

The threshold voltages and hysteresis voltage are given by the following equations:

$$V_{UT} = V_{REF} + \beta(V_{sat} - V_{REF})$$

$$V_{LT} = V_{REF} - \beta(V_{sat} + V_{REF})$$

$$V_H = V_{UT} - V_{LT} = 2\beta V_{sat}$$

V_H is also called the hysteresis window.

The Schmitt trigger circuit is indicated by the symbol ■.

Astable multivibrator (square wave and triangular wave generator)

Unlike bistable multivibrator where both the states are stable, in astable multivibrator none of the two states is stable. They are said to be quasi-stable states. So the output changes state with certain frequency. No external trigger is required to generate the output. Figure 7.17 shows the circuit of an astable multivibrator. This is a modified bistable multivibrator circuit.

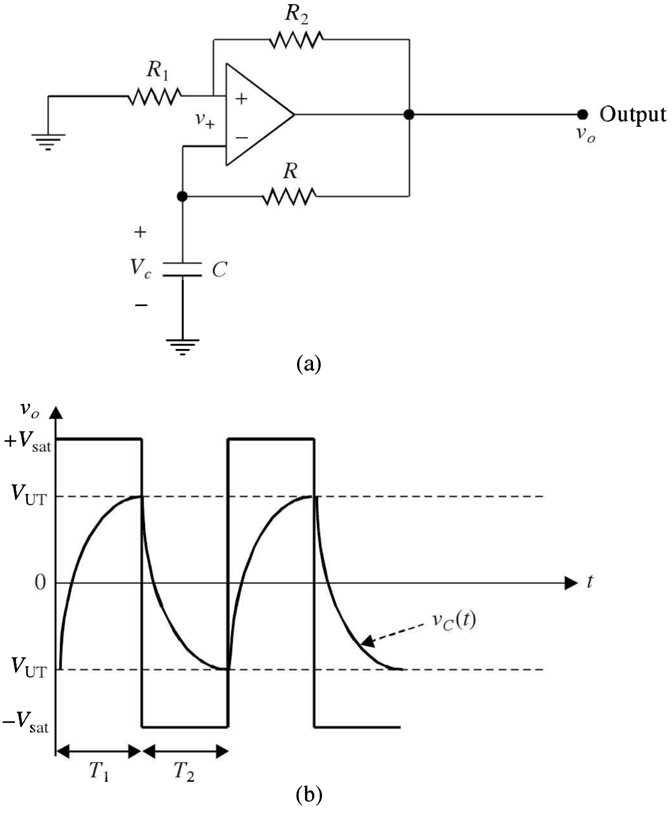


Figure 7.17 (a) Astable multivibrator and (b) signal waveforms across the output and capacitor.

Let us assume that at time $t = 0$, the output voltage is equal to $+V_{sat}$. Capacitor starts charging towards this voltage with time constant equal to RC. As soon as the capacitor voltage equals V_{UT}, i.e. $+\beta V_{sat}$, (say at time T_1) the inverting input voltage of opamp becomes equal to noninverting voltage, v_+. This changes the output to $-V_{sat}$ and capacitor starts discharging towards this negative voltage with the same time constant. When the capacitor voltage equals V_{LT}, i.e. $-\beta V_{sat}$ (say at time T_2) the inverting input voltage becomes equal to noninverting voltage, v_+. This changes the output again to $+V_{sat}$ and capacitor starts charging towards this voltage. The process repeats after time $T = T_1 + T_2$. Figure 7.17(b) illustrates this process. So a square wave is generated at the output for $T_1 = T_2$.

The voltage across the capacitor during charging is

$$v_C(t) = (+V_{\text{sat}}) - [(+V_{\text{sat}}) - \beta(-V_{\text{sat}})]\ e^{-t/RC}$$

At $t = T_1$,

$$v_C(T_1) = +\beta V_{\text{sat}}$$

$$= (+V_{\text{sat}}) - [(+V_{\text{sat}}) - \beta\,(-V_{\text{sat}})]\ e^{-T_1/RC}$$

Therefore,

$$T_1 = RC \ln \frac{1 - \beta[(-V_{\text{sat}})/(+V_{\text{sat}})]}{1-\beta} = RC \ln \frac{1+\beta}{1-\beta}$$

Similarly, we can calculate time T_2 by interchanging $+V_{\text{sat}}$ and $-V_{\text{sat}}$ in the above equation, i.e.

$$T_2 = RC \ln \frac{1 - \beta[(+V_{\text{sat}})/(-V_{\text{sat}})]}{1-\beta} = RC \ln \frac{1+\beta}{1-\beta}$$

The time period of the square wave is

$$T = T_1 + T_2 = 2RC \ln \frac{1+\beta}{1-\beta}$$

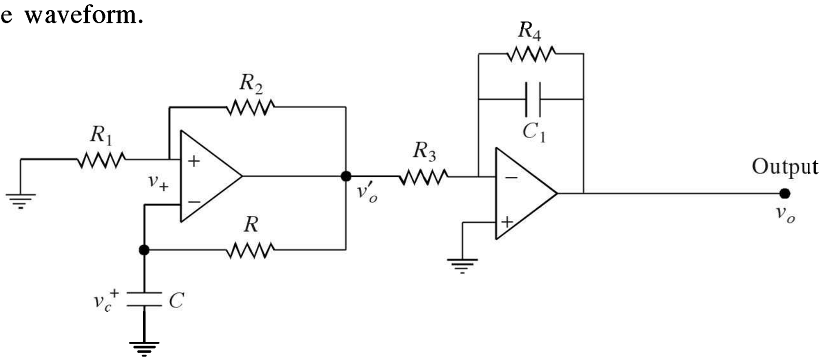
$$= 2RC \ln \left(1 + \frac{2R_1}{R_2}\right) \qquad \text{since } \beta = \frac{R_1}{R_1 + R_2}$$

Note that by changing R and C, we can change the time period and hence the frequency of the square waveform.

Generating triangular waveforms: astable multivibrator

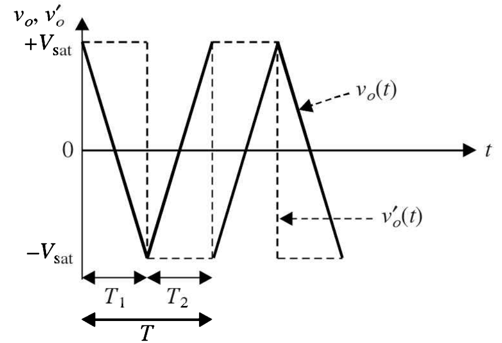
As shown in Figure 7.17, the voltage across the capacitor, v_C, is nearly a triangular wave. To make the slopes more linear, β can be reduced. However, other better circuits are available for generating triangular waveforms.

We can integrate a square wave to get a triangular wave. Figure 7.18 shows the circuit diagram. The output of an astable multivibrator is fed to an integrator. The time constant of the integrator is R_3C_1, assuming R_4 to be large. R_3C_1 is chosen equal to T. Here R_4 is used to stabilize the waveform.



(a)

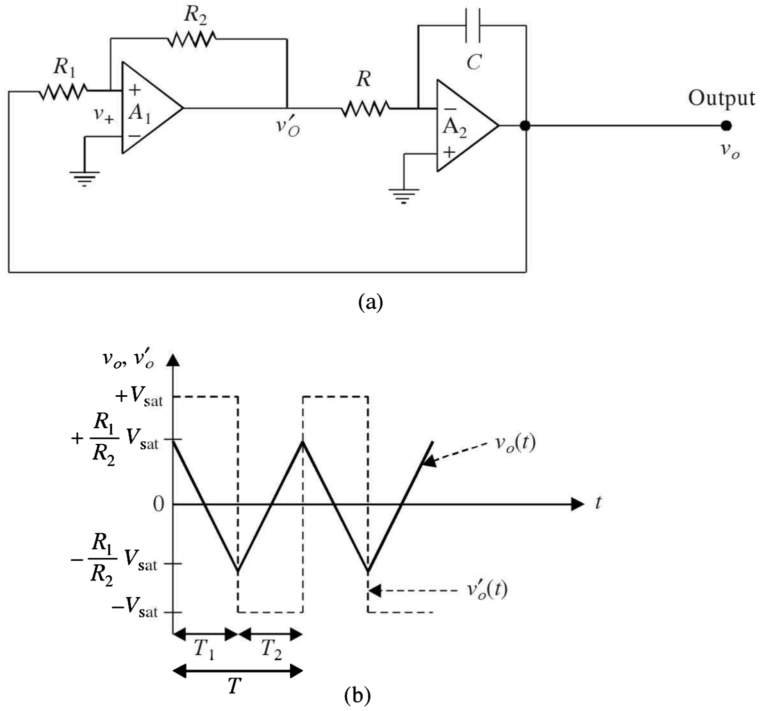
Figure 7.18 (Cont.)



(b)

Figure 7.18 Triangular wave generator: (a) circuit diagram and (b) waveforms.

Another circuit using a bistable multivibrator and an integrator is shown in Figure 7.19. Let us assume that at $t = 0$, $v_o' = +V_{sat}$. This forces the current of $+V_{sat}/R$ through the capacitor and the voltage across the capacitor, v_o, decreases linearly. This continues until v_+ at the noninverting input terminal of A_1 crosses zero and becomes negative. v_o' is then equal to $-V_{sat}$. This reverses the direction of current through the capacitor and v_o increases linearly till $v_o' = +V_{sat}$. The cycle then repeats.



(b)

Figure 7.19 Another triangular wave generator: (a) circuit diagram and (b) waveforms.

At $t = 0$, $v_o' = +V_{sat}$ and current through the capacitor $i_+ = +V_{sat}/R$, and the output voltage, v_o, is given by

$$v_o(t) = -\frac{1}{C}\int_{-\infty}^{t} i_+\,dt = -\frac{i_+ t}{C} + v_o(0)$$

Now the voltage v_+ can be found using superposition

$$v_+(t) = \frac{R_1}{R_1 + R_2} v_o' + \frac{R_2}{R_1 + R_2} v_o$$

Just before $t = T_1$, $v_+ = 0$ and $v_o' = V_{sat}$. Therefore

$$v_o(T_1) = -\frac{R_1}{R_2} V_{sat}$$

Just after $t = T_1$, $v_o' = -V_{sat}$ and v_+ becomes negative. The current through the capacitor between time T_1 and T_2 equals $i_- = -V_{sat}/R$ and v_o starts rising linearly. At $t = T_2$,

$$v_o(T_2) = \frac{R_1}{R_2} V_{sat}$$

and the cycle repeats. Substituting these values in the relationship of $v_o(t)$, at $t = T_1 = T/2$, we get

$$-\frac{R_1}{R_2} V_{sat} = -\frac{T}{2RC} V_{sat} + \frac{R_1}{R_2} V_{sat}$$

This equation gives time period T as

$$T = 4RC\frac{R_1}{R_2}$$

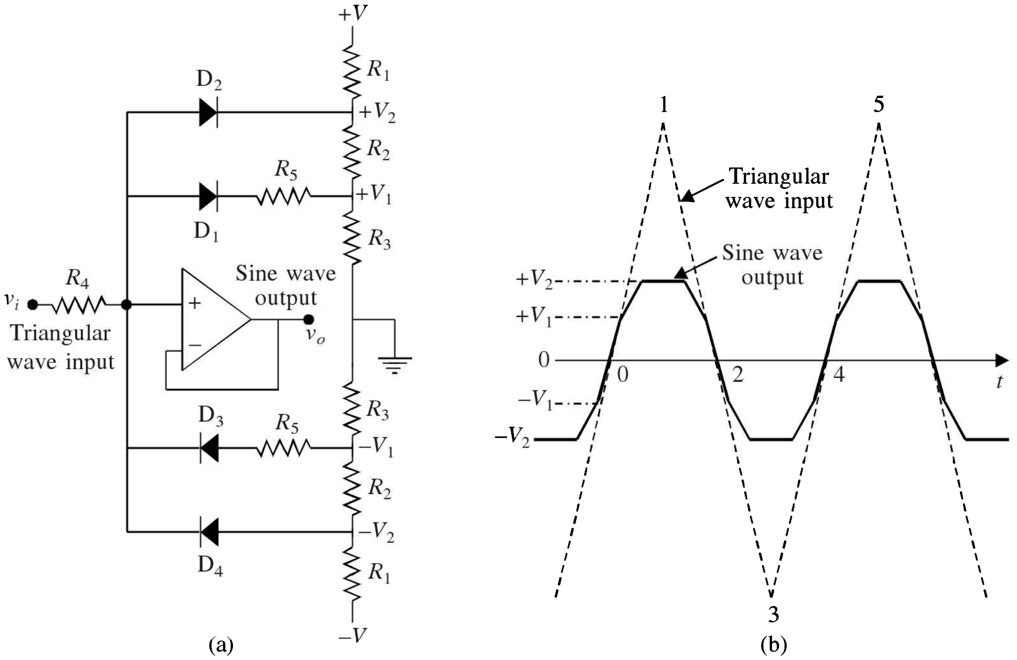
and the frequency, $f = R_2/4RCR_1$. The amplitude of the output = $(R_1/R_2)\, V_{sat}$.

Triangular-to-sinusoidal waveform converter

The most commonly used method to convert a triangular wave into a sine wave is by using a Diode Function Generator. In this method, the input nonlinear curve (here triangular wave) is divided into piecewise linear segments to implement a sine wave. Piecewise linear segments are realized by diodes in series with resistors and dc voltage sources. Figure 7.20 shows a three segment triangular-to-sine wave converter along with the input and output waveforms.

Each quarter is divided into three segments. Consider the first quarter from 0 to 1. When the input voltage is less than V_1, all diodes are reverse biased and the output follows the input. The opamp is being used as a buffer. As the input voltage rises to V_1 or above, diode D_1 conducts. Assuming all diodes to be ideal and also assuming that no input signal current flows through voltage divider resistors, for $v_I > V_1$, we get

$$v_o = V_1 + (v_I - V_1)\frac{R_5}{R_4 + R_5}$$



(a) (b)

Figure 7.20 Triangular-to-sinusoidal wave converter: (a) circuit diagram and (b) waveforms.

which is an equation of a straight line with slope less than one but equal to $R_5/(R_4 + R_5)$. Therefore, the output voltage rises with this slope till v_I is equal to and more than V_2. Then both the diode D_1 and D_2 conduct and since diodes are ideal, the output voltage equals V_2. Now, the slope is zero. This cycle repeats in reverse direction when input goes to the second quarter 1 to 2. Similarly for the third and fourth quarters (2 to 3 and 3 to 4), we get three segments in each quarter but in negative direction (diodes D_3 and D_4 will operate). The complete waveform is shown in Figure 7.20(b).

We observe that the output with three segments is not that pure. The *total harmonic distortion* (THD) is very high. To make the waveform look closer to a sine wave, we should increase the number of segments in each quarter. Six to eight segments can generate a sine wave of very low THD.

Monostable multivibrator (pulse generator)

In this multivibrator one state is stable and the other quasi-stable. A trigger causes the output to make a transition from the stable to the quasi-stable state, where it remains for a predetermined length of time, and then reverts to the stable state. Monostable multivibrators are sometimes referred to as ***one-shots*** since one *shot* from the trigger creates a change in the output. Figure 7.21 shows the circuit and waveforms of a one-shot.

The circuit is in the stable state with $v_o = +V_{sat}$. The voltage at the noninverting input, $v_+ = +\beta V_{sat}$ which is maintained more than the voltage at the inverting input which is equal to cut-in voltage of diode D_1. The voltage, $v_2 = (\beta V_{sat} - V_{D2})$ where V_{D2} is the cut-in voltage of

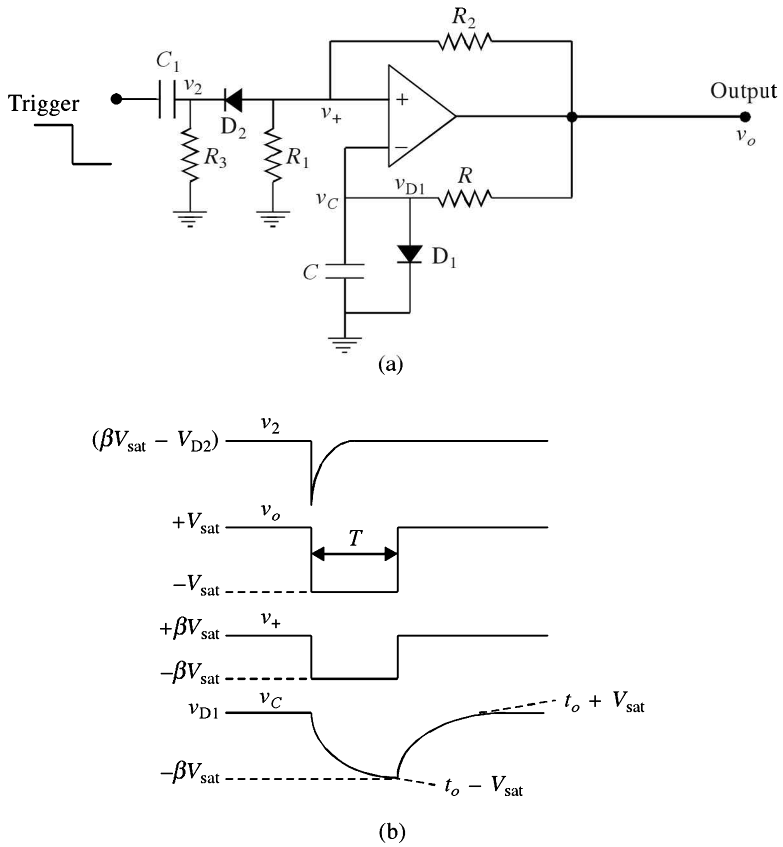


Figure 7.21 Monostable multivibrator: (a) circuit diagram and (b) waveforms.

diode D_1. This state remains unchanged till a negative trigger step voltage is applied at the input. This is differentiated by C_1R_3 giving rise to a negative pulse. If the magnitude of the trigger pulse is sufficient to make $v_+ < v_C$, the output voltage v_o changes to $-V_{sat}$ and v_+ to $-\beta V_{sat}$. This reverse voltage biases diode D_2. Diode D_1 also is reversed biased and capacitor C charges towards $-V_{sat}$. As soon as $v_C = -\beta V_{sat}$, v_o switches back to stable state $+V_{sat}$. The voltage v_C starts building up towards $+V_{sat}$ till it reaches V_{D1}. The output stays in the stable state till another trigger is applied.

The duration of the output pulse, T, is determined from the charging equation of the capacitor C given below:

$$v_C(t) = -V_{sat} - (-V_{sat} - V_{D1})\, e^{-t/RC}$$

At $t = T$, $v_C(t) = -\beta V_{sat}$. Therefore,

$$-\beta V_{sat} = -V_{sat} + (V_{sat} + V_{D1})\, e^{-T/RC}$$

or

$$T = RC \ln \frac{1 + \dfrac{V_{D1}}{V_{sat}}}{1 - \beta} = RC \ln \frac{1}{1 - \beta} \qquad \text{for } V_{D1} << V_{sat}$$

where $\beta = \dfrac{R_1}{R_1 + R_2}$.

7.4 FUNCTION GENERATORS

A function generator is a versatile instrument that most commonly generates square, triangular, sawtooth and sine waves over a wide frequency range. These outputs may be available simultaneously or one at a time depending upon the design of the instrument. Figure 7.22 shows the block diagram of a function generator.

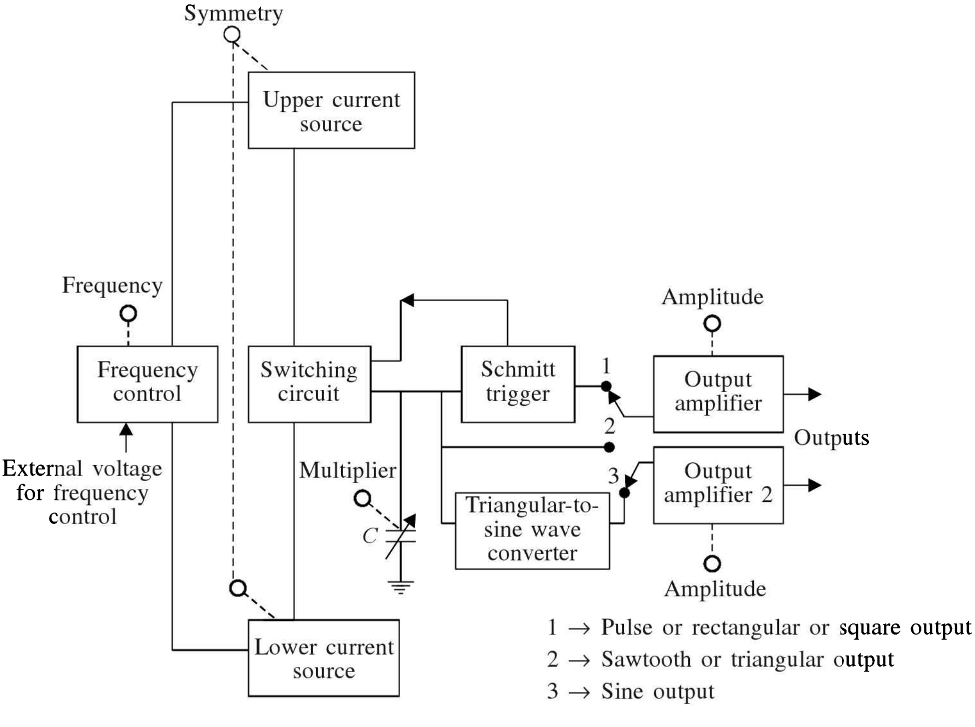


Figure 7.22 Function generator.

There are two constant current sources, upper current source and lower current source which, respectively, charge and discharge the capacitor, *C*. This constant current charging of the capacitor generates two ramps in the form of a triangular or saw-tooth waveform. This can be made available at the output through the amplifier with switch at position 2. This triangular wave is passed through a Schmitt trigger circuit which converts it into a square wave or rectangular wave or pulses. This can be made available at the output through the amplifier with switch at position 1. Note that the Schmitt trigger circuit output is also fed to a switching circuit to facilitate change of current source from upper to lower.

Sine wave is generated by using a triangular wave to sine wave generator. This can be made available at the output through the amplifier with switch at position 3.

Various controls are available from the generator. Frequency Control controls the sum of the currents from the upper and lower current sources by applying control voltages to these sources either from an external source or from the controls available on the Function Generator. Symmetry Control is used to set the ratio of these currents which in turn determines the duty cycle of the output waveform. Multiplier Control changes the capacitor value thereby selecting the size of the ramp. Amplitude of the output is controlled with Amplitude Control.

7.5 PULSE PARAMETER DEFINITIONS

There are many pulse parameters which are used for instrument specifications. Figure 7.23 gives a graphical overview of these parameters. The parameters are explained below:

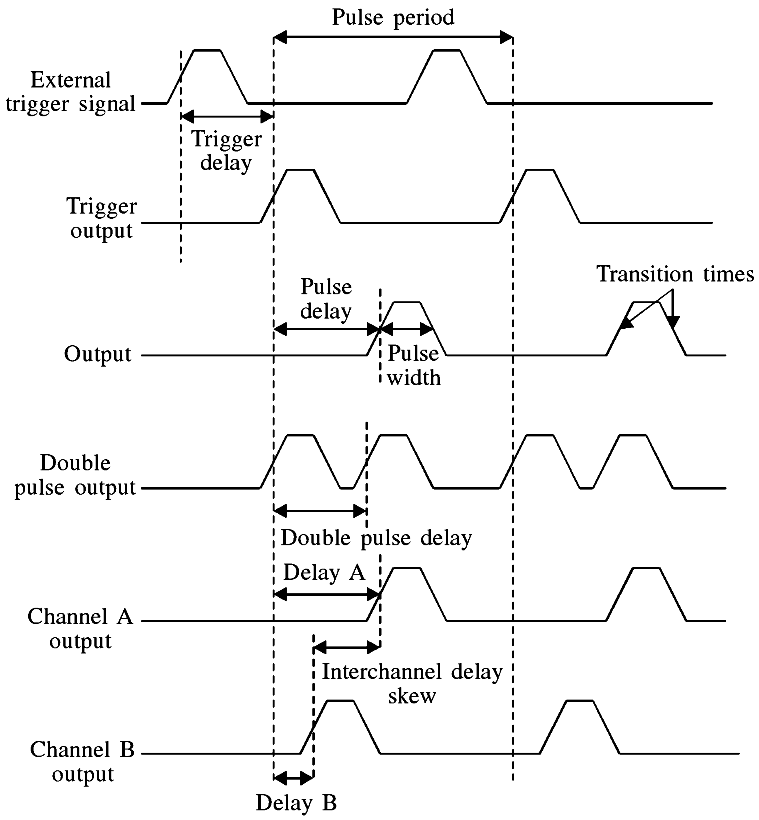


Figure 7.23 Overview of pulse parameters.

Time reference point. This is at the median of the amplitude (50% amplitude point on pulse edge).

Pulse period. The time interval between the leading edge medians of the consecutive output pulses.

Trigger delay. The interval between the trigger point of the external trigger input signal and the trigger output pulse's leading edge median.

Pulse width. The interval between the leading and trailing edge medians.

Pulse delay. The interval between the leading edge medians of trigger output pulse and output pulse.

Double pulse delay. The interval between the leading edge medians of the double pulses.

Interchannel delay (skew). The interval between the corresponding leading edge medians of the output signals.

Transition (rise/fall) time. The interval between the 10% and 90% amplitude points on the leading/trailing edge (see Figure 7.24).

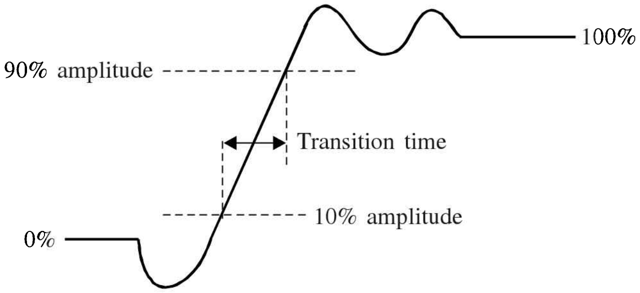


Figure 7.24 Transition time.

Linearity. Peak deviation of an edge from a straight line through the 10% and 90% amplitude points, expressed as percentage of pulse amplitude (see Figure 7.25).

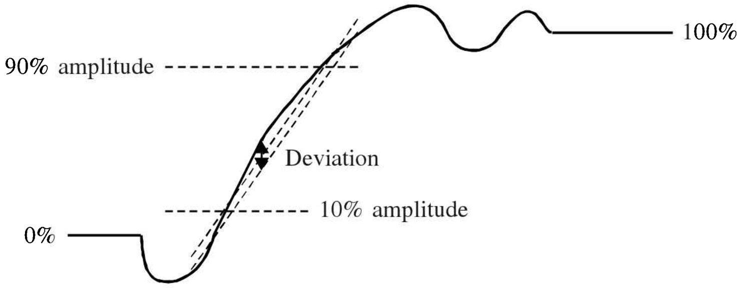


Figure 7.25 Linearity.

Stability. Long-term average instability over a specific time, for example, hour, year. Jitter is excluded.

Pulse levels. Pulse output is specified as pulse top and pulse base (usually referred to as high level and low level), or as peak-to-peak amplitude and median offset. A *window* specification shows the limits within which the pulse can be positioned. (see Figure 7.26).

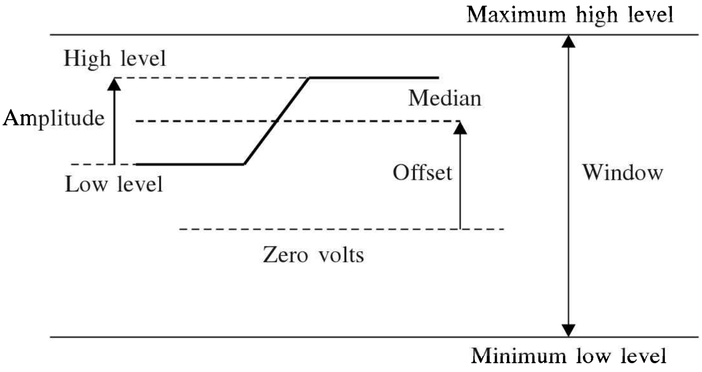


Figure 7.26 Pulse levels.

Preshoot, overshoot, ringing. Preshoot and overshoot are peak distortions preceding/following an edge. Ringing is the positive-peak and negative-peak distortion, excluding overshoot, on pulse top or base. A combined preshoot, overshoot and ringing specification of, for example, 5% implies:

- Overshoot/undershoot < 5%
- Largest pulse-top oscillation < ±5%, of the pulse amplitude.

Figure 7.27 illustrates these definitions.

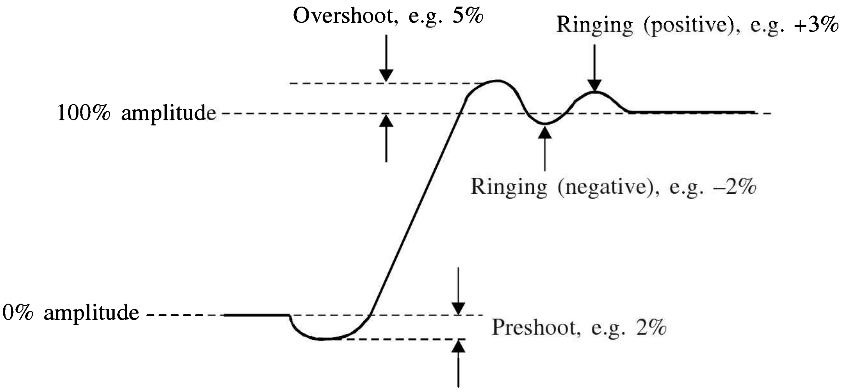


Figure 7.27 Preshoot, overshoot, ringing.

Settling time. Time taken for pulse levels to settle within level specifications, measured from 90% point on leading edge. (see Figure 7.28).

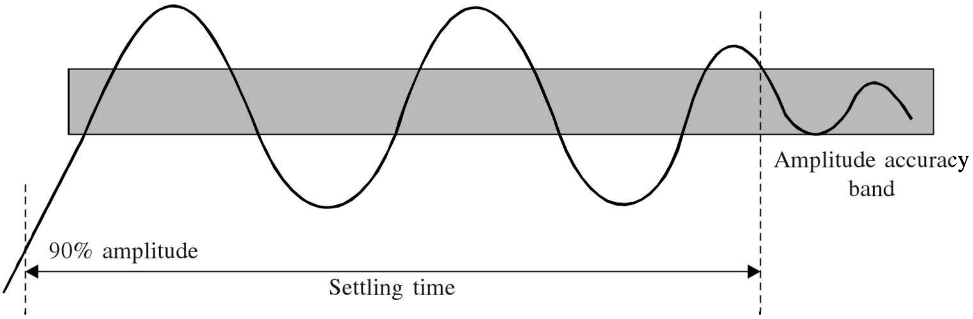


Figure 7.28 Settling time.

Repeatability. When an instrument operates under the same environmental conditions and with the same settings, the value of a parameter will lie within a band inside the accuracy window. Repeatability defines the width of this band (see Figure 7.29).

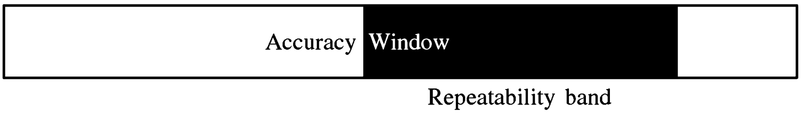


Figure 7.29 Repeatability.

Jitter. This is the short term instability of one edge relative to a reference edge. That is, the variation in timing of the edges (transitions) of a pulse train relative to the timing of a perfect, jitter-free clock. For instance, if a period of 20 ns is preprogrammed, the internal oscillator of a pulse generator may produce one pulse with 20.01ns followed by 19.99ns and so on. Jitter may occur for several reasons:

- The clock may have some amount of intrinsic jitter due to noise in the clock oscillator.
- Noise coupled into a transmission link will instantaneously add to or subtract from the signal amplitude, causing transitions to occur earlier or later.
- Even under noise-free conditions and with perfect, jitter-free clocks, reduced bandwidth (as caused by long digital cables) will introduce jitter.
- The power supply may have inadequate regulation (noise).

Jitter is often specified as rms value, which is one standard deviation, σ. If the distribution is assumed Gaussian, the value of 6σ represents 99.74% of the peak-to-peak jitter. For example, if jitter specification is 20 ps rms, the peak-to-peak jitter should be no worse than 120 ps. If a pulse generator is used as a high speed clock source, low jitter is essential. For example, if a circuit is to be tested at 500 MHz (period = 2000 ps), a jitter specification of 50 ps rms will result in peak-to-peak jitter of 300 ps which is 15% of the period. If jitter specifications are 10 times better, circuit could be tested at a clock rate that is 10% faster, because a much lower noise budget would be required. Figure 7.30 shows the different types of jitters. These are defined with reference to the triggering reference edge, *E*.

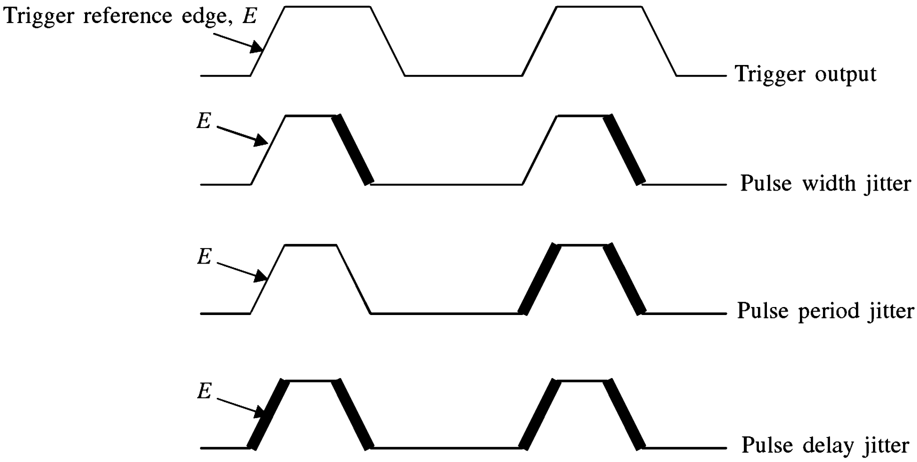


Figure 7.30 Different types of jitters.

- **Width jitter** is stability of trailing edge with reference to leading edge, *E*.
- **Period jitter** is the stability of a pulse with reference to previous leading edge.
- **Delay jitter** is the stability of a pulse with reference to leading edge of the trigger output.

7.6 FREQUENCY SYNTHESIS TECHNIQUES

A frequency synthesizer is an electronic device which generates one or more frequencies from a reference source and some control signals. The main architectures for frequency synthesizers are:

1. **Direct analog synthesis** (also called mixer-filter-divider), in which the output frequency is obtained directly from the reference frequency by the operation of mixing, filtering, multiplication, and division.
2. **Indirect synthesis** (phase locked loop) in which the output frequency is obtained from a secondary oscillator (usually a VCO) that is phase locked (sometimes only frequency locked) to the primary frequency reference.
3. **Direct digital synthesis** (DDS) which generates digitally the output frequency from the reference frequency.

The figures of merit that can be considered in selection of a synthesizer are:

(i) Spectral purity (level of output spurious component and noise with respect to an ideal output)

(ii) Tuning bandwidth (output frequency range)

(iii) Tuning (or switching) speed

(iv) Frequency resolution

(v) Number of generated frequencies
(vi) Flexibility (possibility of different modulations)
(vii) Phase continuity

These aspects are tied together and each of the three methods has advantages and disadvantages. For example, in a frequency hopping spread spectrum system, the synthesizer must be capable of tuning quickly, in small frequency steps, over a wide band. Wide tuning bandwidth is best provided with indirect synthesis but better switching speed makes the DDS the best suited for use in a Spread Spectrum System. Direct Digital Synthesis is the technique that has had the best improvements in the last few years. In fact the idea behind it is quite elementary. But one problem for first DDS realizations was the high number of devices necessary for its implementation and its low speed that resulted in small bandwidth; the current developments in integration of electronic circuits permit the reduction of space used and an increase in reliability.

7.6.1 Direct Analog Synthesis

In this technique, multiple frequencies are simultaneously generated from a common reference. which are then selected and assembled in various combinations to produce the desired frequency at the output. Figures 7.31 and 7.32 show two examples to generate signals in the range 1–9 MHz and 1–99 MHz in increments of 1 MHz. BPF in the figure is a band-

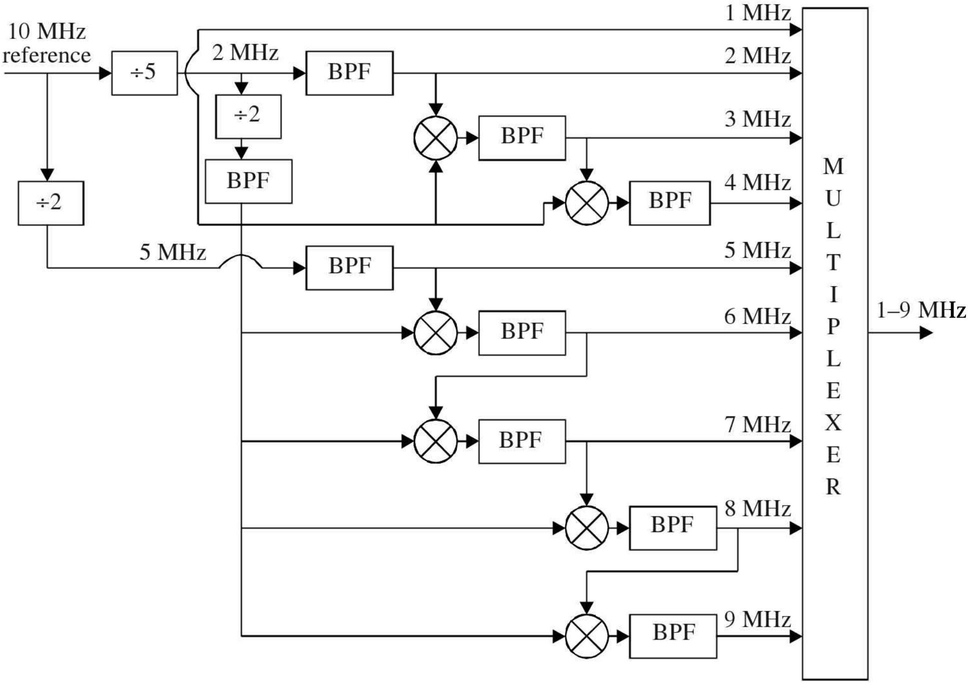


Figure 7.31 Direct analog synthesizer in the range 1–9 MHz.

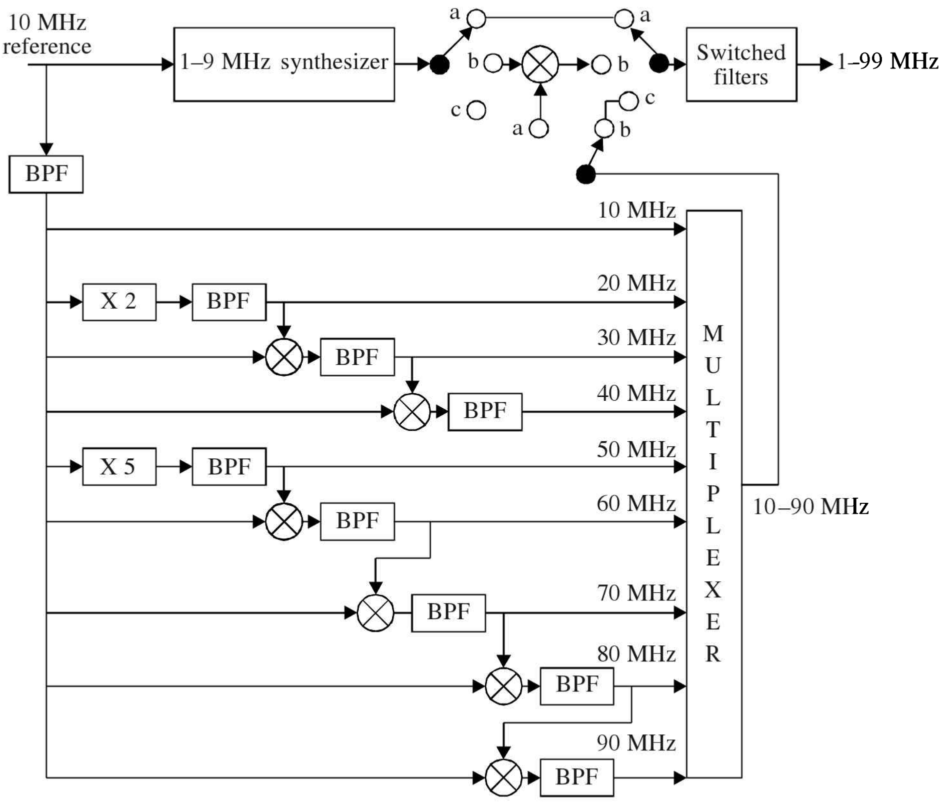


Figure 7.32 Direct analog synthesizer in the range 1–99 MHz.

pass filter. A more agile (fast switching) synthesizer using the mix and divide technique, to provide decades of fine frequency control in a modular architecture, is shown in Figure 7.33.

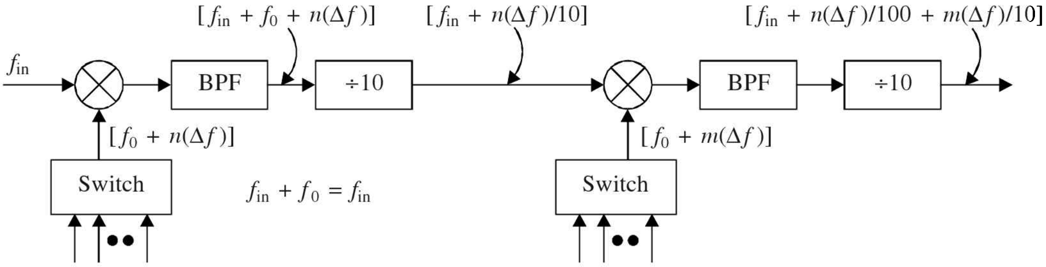

$[f_{in} + f_0 + n(\Delta f)]$
$[f_{in} + n(\Delta f)/10]$
$[f_{in} + n(\Delta f)/100 + m(\Delta f)/10]$
f_{in}
BPF
÷10
BPF
÷10
$[f_0 + n(\Delta f)]$
$[f_0 + m(\Delta f)]$
Switch
$f_{in} + f_0 = f_{in}$
Switch

Figure 7.33 Fast switching synthesizer using mix and divide technique.

Each identical module contains a mixer, a switch selecting one of the ten input frequencies, a band-pass filter which selects upper side band from the mixer, and a ÷ 10 divider. The same 10 frequencies are supplied to each module and satisfy the relationship

$$f_k = f_0 + k(\Delta f)$$

where k is an integer $1 \le k \le 10$, Δf is the constant interval between each of the input frequencies, and the frequency f_0 is chosen such that

$$f_{in} + f_0 = 10 f_{in}$$

Thus the output frequency selected by the band-pass filter after the first mixer is

$$f_{in} + f_0 + n(\Delta f) = 10 f_{in} + n(\Delta f)$$

The output frequency after the first divider is $f_{in} + n(\Delta f)/10$. When this frequency is fed to the second modular stage, its output frequency becomes

$$f_{out} = f_{in} + \frac{n(\Delta f)}{100} + \frac{m(\Delta f)}{10}$$

This illustrates how decades of frequency step size can be realized in a modular fashion. For example, for $f_{in} = 1$ MHz, $f_0 = 9$ MHz, $\Delta f = 1$ MHz, $n = 0$ to 9 and $m = 0$ to 9, the output frequency will vary from 1.00 to 1.99 MHz.

The principal advantage of direct analog synthesis technique is the speed with which the output frequency can be changed. But it is hardware-intensive (therefore expensive and more power consuming) and phase continuity may be lost during switching. It is also very much prone to spurious (both harmonics and nonharmonics) signals in the output because of the multiplicity of mixers.

7.6.2 Indirect Synthesis

Indirect frequency synthesizers employ a secondary oscillator controlled by a phase-locked loop (PLL) or a frequency-locked loop to generate the output frequency as shown in Figure 7.34. This type of synthesizer is relatively simple and inexpensive compared to a direct synthesizer. Because of their small-size and low power consumption, miniature radio transceivers rely on phase-locked loop synthesizers. The basic form of the PLL consists of a voltage-controlled oscillator (VCO), a frequency divider, a phase detector, and a loop (low-pass) filter.

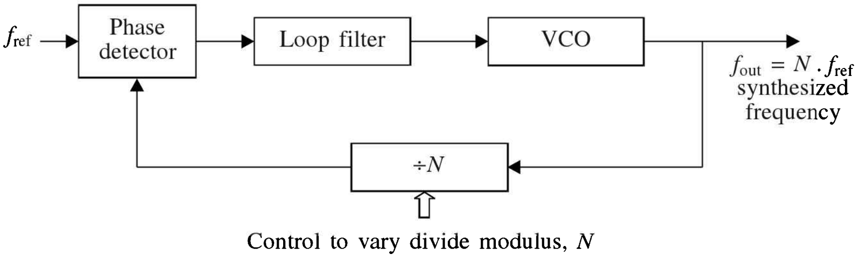

f_{ref}
Phase detector
Loop filter
VCO
$f_{out} = N \cdot f_{ref}$ synthesized frequency
÷N
Control to vary divide modulus, N

Figure 7.34 Block diagram of multiplier phase-locked loop.

The VCO generates the synthesized frequency according to a dc level at its input. Such an arrangement allows the possibility of slaving a synthesized frequency to a spectrally pure, lower reference frequency. In the lock condition, the VCO output frequency f_{out} is related to the reference frequency f_{ref} by $f_{\text{out}} = Nf_{\text{ref}}$, where N is the divider modulus. This indicates that the smallest frequency increment generated by the loop is equal to the reference frequency.

For example, if the reference frequency is 1 MHz and N is 10, then the output frequency resolution is 1 MHz and the output frequency is 10 MHz. If we want resolution of say 1 Hz, we need to use a reference frequency of 1 Hz. For the same frequency output (10 MHz), the value of N should be 10^7.

The switching speed of the PLL is a function of the loop bandwidth and closed-loop frequency response. Since the loop bandwidth is always smaller (by about 10 times) than f_{ref}, in order to provide a stable correction signal to the VCO, faster switching speeds require higher reference frequencies and therefore higher loop bandwidths. However, the loop design parameters cannot be chosen to optimize all aspects of circuit performance. The result is a trade-off between narrow bandwidths for maximum filtering and wide bandwidths to provide short rise-times for fast tracking of the input signal. Further, the frequency multiplier loops also multiply noise at the phase detector, which appears as noise sidebands on the oscillator output.

Multiple-loop indirect synthesizer

When fine resolution is needed, sequences of multiplication, division and addition are used, involving a number of loops. Figure 7.35 shows an example of a multiple-loop indirect synthesizer to synthesis 15.1 MHz frequency from a 10 MHz reference source.

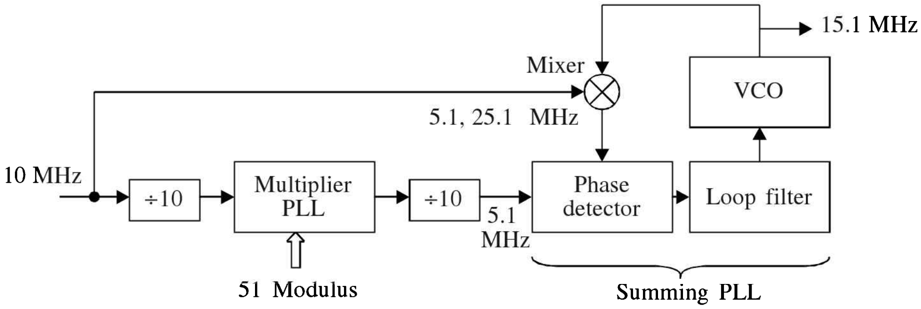


Figure 7.35 A multiple-loop indirect synthesizer.

Note that the input to summing PLL is 5.1 MHz. It adds 10 MHz into this input to generate 15.1 MHz.

Techniques like this can be used to get any desired resolution. Since multiplication numbers are low and the loop reference frequency is high, the output will have low noise sidebands and be capable of fast switching. All this occurs at the cost of greater complexity.

Programmable decade indirect synthesizer

This technique is used to synthesize frequencies in decimal points. Figure 7.35 shows one scheme for two decimal points. It uses a high reference frequency allowing wide loop bandwidth and thereby giving faster speed. However, the hardware requirement increases with increase in decimal points.

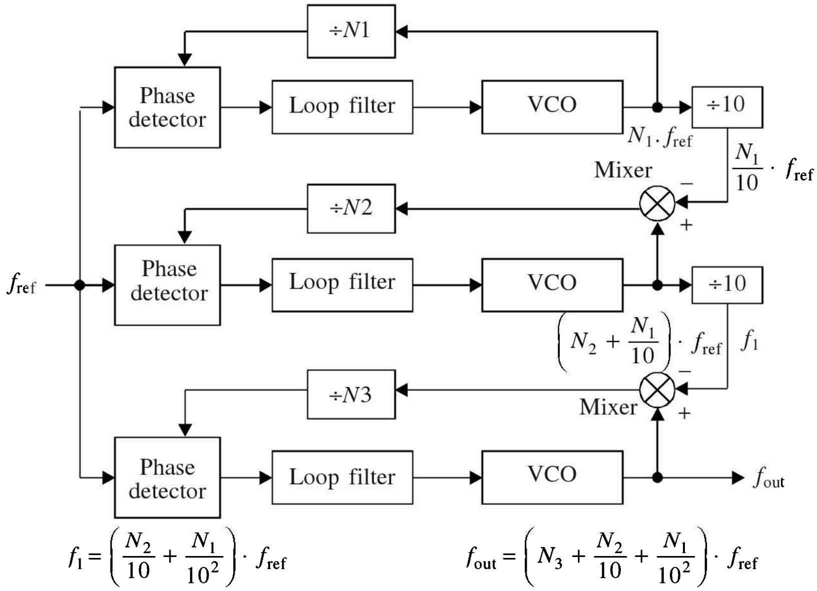

÷N1
Phase detector
Loop filter
VCO
÷10
$N_1 \cdot f_{ref}$
Mixer
$\frac{N_1}{10} \cdot f_{ref}$
÷N2
Phase detector
Loop filter
VCO
÷10
f_{ref}
$\left(N_2 + \frac{N_1}{10}\right) \cdot f_{ref}$
f_1
÷N3
Mixer
Phase detector
Loop filter
VCO
f_{out}
$f_1 = \left(\frac{N_2}{10} + \frac{N_1}{10^2}\right) \cdot f_{ref}$
$f_{out} = \left(N_3 + \frac{N_2}{10} + \frac{N_1}{10^2}\right) \cdot f_{ref}$

Figure 7.36 Programmable decade indirect synthesizer.

Fractional—N indirect synthesizer

As pointed out earlier, the response time of the PLL depends largely on the loop bandwidth. Consequently, to obtain fast switching it is necessary to reduce the loop response time by widening the loop bandwidth. One possible way of achieving a fine frequency step without either lowering the reference frequency or considerably increasing the divider ratio N, is to make use of a fractional division approach. The basic fractional-N phase-locked loop technique involves use of a high-speed frequency divider working in conjunction with a low-speed programmable counter whose division ratio is the sum of an integer and a fraction of an integer. The fractional division is substituted for the fixed programmable divider of conventional PLL synthesizers. This approach is also known as *variable-modulus dividing* or the *pulse swallowing technique*.

Standard frequency synthesis with a swallow counter

Before discussing fractional-N synthesis, the related technique of PLL frequency synthesis utilizing a swallow counter will be discussed as a way to introduce the concepts involved. This method of frequency synthesis provides the advantage of using a single high-speed, dual-modulus prescaler (with division ratio selectable to be either P or $P + 1$), followed by a slower counter whose output connects to the phase detector.

The structure of the frequency synthesizer is shown in Figure 7.37. The dual-modulus prescaler has basically two counters in parallel whose division ratio varies by one (for example, ÷2 or ÷3) followed by a single fixed length counter (for example, ÷64). The swallow enable control allows the higher value counter (÷3) to reach its full count once, and

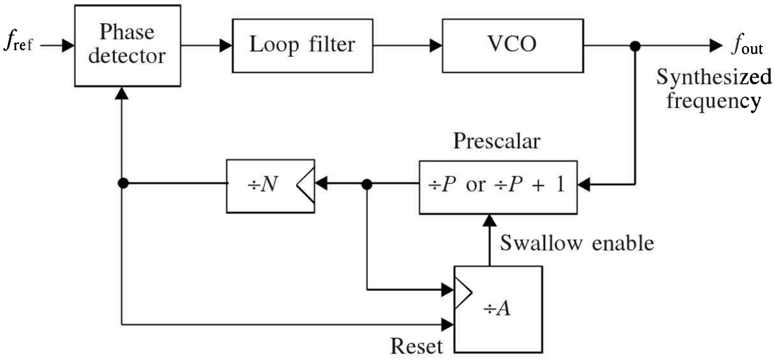


Figure 7.37 Swallow counter frequency synthesizer.

advance the fixed counter from a 0 to 1. Once this occurs the lower value counter (÷2) is used to clock the fixed divider until the fixed counter reaches its full count (63). The process repeats until the swallow enable is set to logic LOW. The net effect of activating the swallow enable control is to increase the prescaler divide ratio from its base value P to $P + 1$ by *swallowing* one VCO pulse.

An additional counter clocked by the prescaler output (here referred to as the A counter) is added to the PLL structure to provide the required enable control signal for the swallow counter. The low speed counter connecting the output of the prescaler to the phase detector (here referred to as the N counter) must be programmed to any value greater than A. Otherwise, the prescaler division ratio will always be $P + 1$. Initially, the prescaler's division ratio is $P + 1$ as both counters begin counting down from their initial values. When the A counter reaches zero, it is disabled and the prescaler division ratio changes to P. The N counter continues to count down. Once the N counter reaches zero, both counters are reloaded with their initial values, and the process is repeated.

The operation of the swallow counter frequency synthesizer can be compared to a weighted averaging process. The average frequency appearing at the output of the N counter is the sum of A prescaler output cycles, counted with a division ratio of $P + 1$, and N-A cycles, counted with a division ratio of P, averaged over one reference cycle. In this case, the average reduces to a simple sum; however, the fractional-N synthesis methods average the divider output frequency over more than one reference cycle making the concept of the frequency average an important one. Performing the sum indicated previously, the resulting relationship between the reference frequency and the VCO frequency is given by

$$f_{\text{out}} = f_{\text{ref}} \cdot [(N - A) \cdot P + A \cdot (P + 1)]$$

or

$$f_{\text{out}} = f_{\text{ref}} \cdot (N \cdot P + A)$$

The above equation shows that selecting the values of A and N involves a potentially hidden consideration. Since the maximum value of the A counter should be at least equal to the number of channels required, the N counter must be at least one plus this number. The reference frequency required for a small frequency step size using this method must be very low, raising problems associated with PLL lock time and reference spur suppression.

Additionally, many of the commercially available PLL ICs use a higher reference frequency connected to a frequency divider as the reference source applied to the phase detector. In effect, this scheme still utilizes a low frequency reference, but it also provides some minor advantages in terms of phase noise performance.

Standard fractional-N synthesis

Standard fractional-N frequency synthesis extends the technique described above to generate output frequencies at non-integer multiples of the reference oscillator frequency. The traditional architecture of the fractional-N PLL synthesizer is illustrated in Figure 7.38.

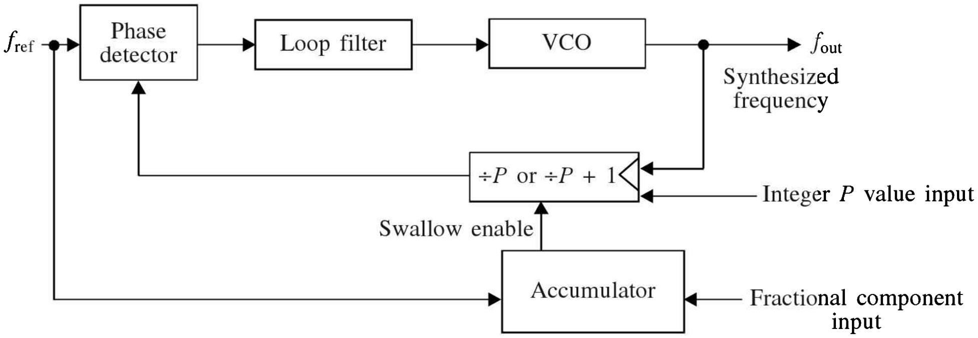


Figure 7.38 Standard fractional-N synthesis.

The structure of the loop is similar to the frequency synthesis using a swallow counter method, previously discussed, with a few significant changes. The value of the prescaler division ratio, P, is externally programmable to provide the coarse tuning of the frequency synthesizer (integer steps of the reference frequency). The division ratio of the loop is also programmable in non-integer steps of the reference frequency by programming the accumulator block.

To understand how the fractional tuning is achieved, the operation of the accumulator should be understood. The accumulator consists of a multi-stage, fixed-length, binary adder and two registers as illustrated in Figure 7.39. One register is programmed with a binary

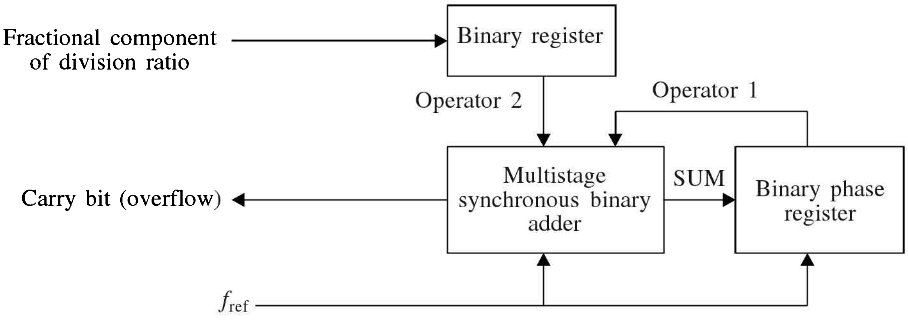


Figure 7.39 Example of accumulator block.

number related to the fractional portion of the relationship between the reference frequency and the output frequency. The binary number contained in the register is added to the contents of other binary phase register. The result of the binary addition is then stored in the binary phase register. The circuit arrangement essentially creates a counter whose count increment is programmable instead of being fixed at unity. The carry output (also termed the *overflow* output) of the adder controls when the division ratio is changed from P to $P + 1$. Since the accumulator advances every reference cycle, the prescaler division ratio is only $P + 1$ for one reference cycle. If the value contained in the binary register (operator 2) is denoted as A, and the minimum sum that causes the adder to overflow is N, then a carry will occur every N/A reference oscillator cycles. For example, if A is 1 and N is 10, it would take exactly ten reference cycles for the accumulator to generate a carry. After N reference oscillator cycles, A pulses of the VCO will have been divided by $P + 1$ and N-A pulses will have been divided by P. Thus, the VCO frequency maintains an integer relationship with the reference oscillator frequency divided by N, which is termed the *base frequency*. The relationship between the VCO frequency and the base frequency is then the same as in the above equation, with the exception that the reference frequency must be replaced by the base frequency. The expression can then be rewritten to describe the relationship between the reference frequency and the VCO frequency as

$$f_{\text{out}} = (f_{\text{ref}} \div N) \cdot (N \cdot P + A) = f_{\text{ref}} \cdot (P + A/N)$$

Thus P will decide the integer and A and N can be chosen to fix the fraction part of the multiplier.

As a result of the integer relationship between the base frequency and the VCO frequency, spurious signals appear at offsets that are integer multiples of the base frequency from the VCO frequency.

The advantage of the standard fractional-N method is the use of a higher reference frequency. Additionally, the higher reference frequency implies a faster acquisition than could be afforded with a very low frequency reference. Finally, the lower overall division ratio means that a lower reference oscillator phase noise contribution will be present on the VCO signal.

7.6.3 Direct Digital Synthesis

Direct-digital synthesis (DDS) is the most recently developed frequency synthesis technique. All three techniques have been available to designers for decades, but it is the DDS that is evolving most rapidly today. In fact, only a few years ago the DDS was a phenomenon with little utility, but now it is an important design tool that cannot be ignored by the architects of any system demanding frequency agility. All other signal generation techniques begin with some sort of oscillator, the output of which is manipulated or controlled by the synthesizer. The DDS is unique because it is digitally deterministic; the signal it generates is synthesized from a digital definition of the desired result.

A DDS uses logic and memory to digitally construct the desired output signal, and a data conversion device to convert it from the digital to the analog domain. Therefore, the DDS method of constructing a signal is almost entirely digital. The phase, the amplitude and the frequency are exactly known at each instant and there are no problems of time or temperature drift. Problems emerge when this digital frequency must be converted in the analog world. In fact, the digital-to-analog converter is one of the critical component in a DDS design. All kinds of modulations are easily implemented with a DDS. Flexibility in the output waveform is also obtained by the fact that it is usually determined by the content of a memory (that can be changed easily). Phase continuous frequency changes are also easily accomplished, while it is not possible to have them in the direct analog synthesis and in a PLL synthesize; they are only possible for very small frequency changes.

DDS architecture provides exceptional frequency agility, making it easy to program both frequency and phase changes on the fly. Agility is the ability to change quickly and cleanly from one frequency to another. It is most useful for testing any type of FM UUT (Frequency Modulated Unit Under Test) device—radio and satellite system components. If the specific DDS's frequency range is sufficient, it is an ideal signal source for tests on FSK (Frequency Shift Keying) and frequency hopping telephoning technologies such as GSM (Global System for Mobile Communications).

Principle of DDS operation

The DDS computes periodically, in real time, a linearly increasing phase angle and then translates it to amplitude values; these samples are digital-to-analog converted and smoothed in time by a filter to produce analog frequency signals. The DDS uses a single frequency to establish a stable sampling time at which the sample values are computed. There are different ways of calculating the amplitude, and different waveforms can be obtained. It is possible to use a ROM as look-up table, algorithmic techniques or simply a kind of inverter if a triangular wave is to be generated. Figure 7.40 shows the block diagram of a DDS.

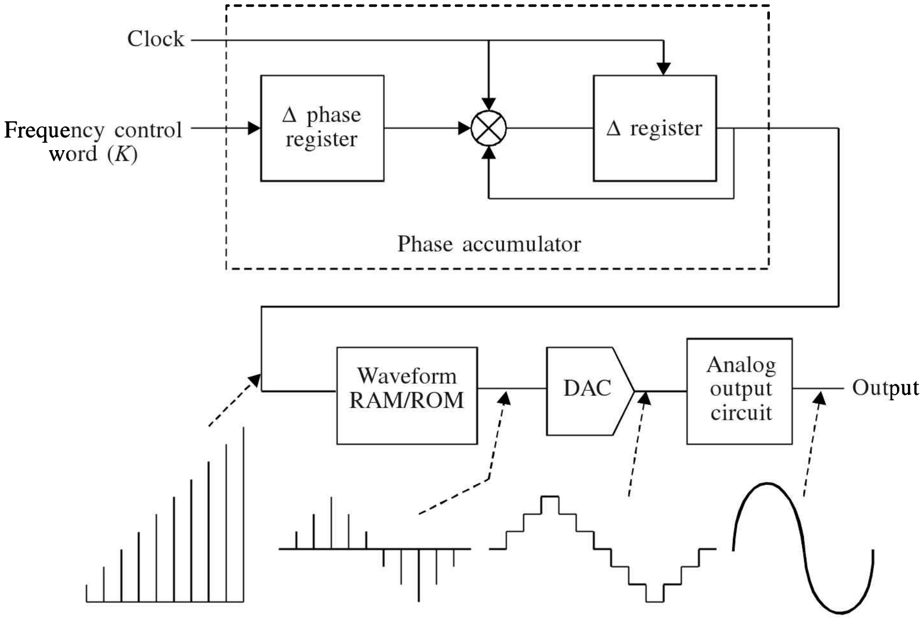


Figure 7.40 Block diagram of a DDS.

In the phase accumulator circuit, the Δ phase register receives instructions from a frequency controller, expressing the phase increments by which the output signal will advance in each successive clock cycle. In modern DDS, the phase resolution may be as small as one part in 2^{30}, that is, approximately 1/1,000,000,000.

The output of the phase accumulator serves as the input for the waveform memory which typically contains just a few basic signals such as sine and square waves. DAC converts this digital signal to analog waveform. The analog output circuit is basically a fixed-frequency low-pass filter which ensures that only the programmed frequency of interest (and no clock artifacts) leaves the output.

In a DDS scheme, sine wave is a typical signal to be generated. Its equation is

$$x(t) = \sin [2\pi f_{\text{out}} t + \phi(t)]$$

The DDS generates samples at determined *instants*, $t = n.t_{\text{ref}}$, where t_{ref} is the sampling interval, $f_{\text{ref}} = 1/t_{\text{ref}}$ is the reference frequency, $\phi(t)$ is phase noise (phase fluctuation) and $n = 0, 1, \ldots$. Phase fluctuation $(\phi(t))$ and frequency fluctuation are related to each other. Frequency fluctuation = $(1/2\pi) \,.\, d\phi(t)/dt$.

Each amplitude sample, $x(nt_{\text{ref}})$, is calculated exploiting a phase sample $\phi(n) = 2\pi f_{\text{out}}\, nt_{\text{ref}}$ with $f_{\text{out}} = Kf_{\text{res}}$ where f_{res} is the frequency resolution and also the minimum frequency that can be generated and is equal to $f_{\text{ref}}/2^N$ where N is the number of bits of the accumulator.

The value of the possible output frequencies is given by the following equation, which is called the DDS tuning equation:

$$f_{\text{out}} = \frac{K f_{\text{ref}}}{2^N}$$

For $\phi(t) = 0$ (no phase noise) and $t = nt_{\text{ref}}$, we have

$$\begin{aligned} x(nt_{\text{ref}}) &= \sin (2\pi f_{\text{out}} nt_{\text{ref}}) \\ &= \sin \left(2\pi \frac{K f_{\text{ref}}}{2^N}\, nt_{\text{ref}} \right) \\ &= \sin \left(2\pi \frac{Kn}{2^N} \right) \\ &= \sin \left(\frac{2\pi\theta(n)}{2^N} \right) \end{aligned}$$

This series of samples is dependent on n and K; n can be seen as a time index and K as a frequency index. With K fixed and n varying it is possible to obtain the addresses for the samples of the sine for one determined frequency. If K is different, for the same value of n, a different sample is obtained corresponding to a different frequency. $\theta(n)$ is the sequence generated by the accumulator and is equal to $<n.K>$ where $<.>$ means modulo 2^N operation.

The accumulator operates on the principle of overflow arithmetic, using the modulo 2^N property of an N bit periodically overflowing accumulator register to simulate the modulo 2π property of the sine function. It is actually made by an N-bit adder followed by an N-bit register. It adds continuously the same value K (also called *frequency setting* or control word, FSW); when it is in overflow (that is, the sum is greater than 2^N-1) it cuts off the most significant bit (that is equivalent to the subtraction of 2^N and takes the maximum integer of the result) and goes on adding. The accumulator thus acts as a digital integrator followed by a modulo 2^N operator. To understand the operation of this system, we can also consider a sine wave oscillation as a vector rotating around a phase circle as shown in Figure 7.41. Each point on the phase circle corresponds to a particular point on the output waveform, and these points are generated as the vector travels around this circle. One revolution on the phase circle corresponds to one cycle of the sine wave. The number of discrete points on the phase circle is determined by the resolution of the phase accumulator. The FSW K represents the *jump size* between updates.

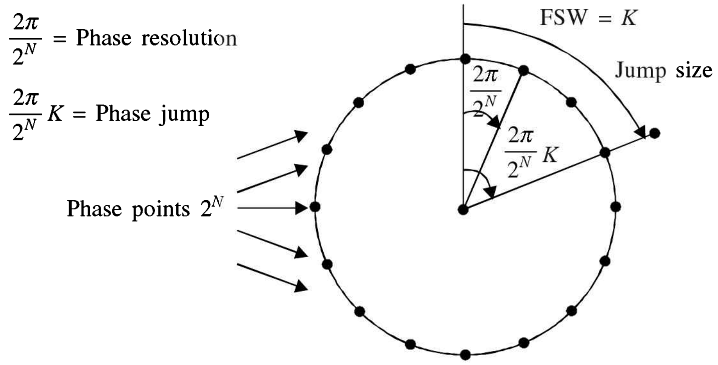


Figure 7.41 Phase circle.

Let us assume that $K = 1$ and Δ phase register is of 30 bits. The accumulator's Δ output register will advance by $360 \div 2^{30}$ in each cycle, since 360° represents a full cycle of the waveform. Therefore, the Δ phase register with value of $K = 1$ produces the lowest frequency waveform requiring the full 2^{30} increments to create one cycle. The circuit will remain at this frequency until a new value of K is loaded into the Δ phase register.

Values of K greater than one will advance through the 360° more quickly, producing a higher frequency (some DDSs use a different approach: they increase the output frequency by skipping some samples, thereby reading the memory contents faster). Note that the only thing that changes is the phase value supplied by the frequency controller. The main clock frequency does not need to change at all. In addition, it allows a waveform to commence from any point in the waveform cycle.

The accumulator output sequence $\theta(n)$ can be seen as the samples of an idealized sawtooth waveform, where the slope is given by $\Delta\theta(n)/\Delta t = K$. It means that the phase accumulator outputs the samples of a hypothetical continuous time sawtooth waveform of amplitude 2^N and period $2^N/K$, (see Figure 7.42). Since the sine function is calculated from the phase sequence, the period of the synthesized output waveform is also $2^N/K$.

The values from the accumulator must be translated to amplitude samples and this can be done in different ways as follows:

(i) It is possible to calculate the value of the waveform in the point in some way (for example with a suitable polynomial approximation or algorithmic techniques).

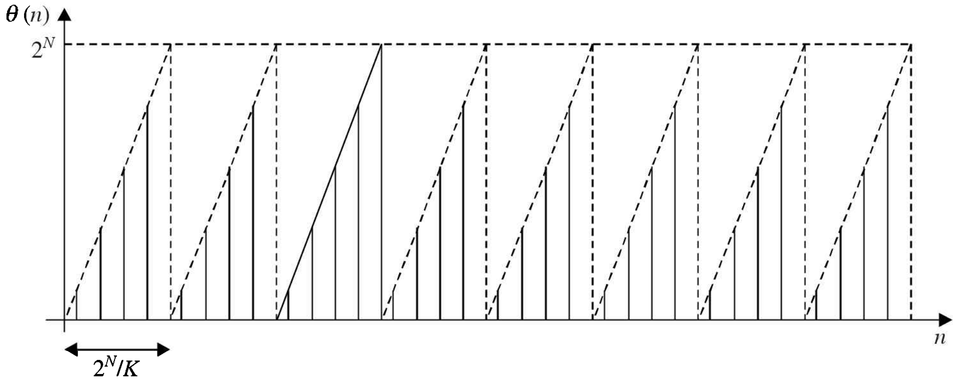


Figure 7.42 Output phase sequence as sampled sawtooth.

(ii) Another method is to use a ROM like a look-up table;

(iii) If a triangular wave can be accepted as the output of the DAC, a simple inverter can be used.

The first approach is the most spontaneous one but the requirement of increased speed has made the ROM-DDS the most popular method. In this approach the computation is minimized at the expense of stored values. In fact, the values of the amplitude for each phase are calculated and stored in the ROM and then they are read. Also the decreasing cost of the ROM has pushed usage in this direction. But now the increasing necessity of speed and space optimization has made the combination of look-up table and computation the most followed. Some of the most recent DDSs are implemented with processors, which calculate the amplitude of the samples with an algorithm. This approach permits to have a phase accumulator of more than 30 bits.

The frequency (and therefore phase) resolution depends on N, the number of bits of the accumulator ($f_{\text{res}} = f_{\text{ref}}/2^N$ and $\phi_{\text{res}} = 2\pi/2^N$) and the amplitude resolution depends on the number of output bits of the ROM. If all the bits from the accumulator are fed into the ROM, then to obtain a very high frequency resolution, a large ROM is needed. This effect is stronger for the phase resolution than for the amplitude resolution. This is because the hardware complexity is a linear function of the word-length of both the ROM output and DAC, while it is an exponential function of the ROM address length. Large ROMs give good output frequency resolution, but limit the speed of the synthesizers (and therefore the bandwidth) because of their excessive access time. Another drawback is that semi-custom approach is not very convenient to implement a large ROM. For these reasons, usually, not all the bits from the accumulator are fed to the input of the ROM, but only a certain number M, so that the $B = N - M$, thus less significant bits from the accumulator are truncated. This is called *phase truncation.* This advantage in the ROM size is paved with differences between the output waveform and the ideal sine; actually there is a loss of information, and so the presence of spurious components in the output spectrum.

In addition to the truncation of the less significant bits from the accumulator, many ways of storage and compression to reduce the ROM size (mixing the look-up table and the computational method) have been developed.

DDS advantages

The advantages of DDS technique are:

- Micro-hertz tuning resolution of the output frequency and sub-degree phase tuning capability, all under complete digital control, can be achieved.
- Extremely fast *hopping speed* in tuning output frequency (or phase), phase-continuous frequency hops with no over/undershoot or analog-related loop settling time anomalies.
- The DDS digital architecture eliminates the need for the manual system tuning and tweaking associated with component aging and temperature drift in analog synthesizers.
- The digital control interface of the DDS architecture facilitates an environment where systems can be remotely controlled, and minutely optimized, under processor control.

DDS limitations

Being a sampled data system, the DDS technique is subject to inherent limitations of such systems:

Quantization noise. This is caused by finite word length. Usually enough resolution is provided to keep noise comparable to other synthesis techniques.

Aliasing. Aliasing takes place when a waveform is under sampled. This produces waveform of much lower frequency than the actual frequency. For example, if the maximum input frequency is 70 Hz and the sampling frequency is 100 Hz, it will produce an alias frequency of (100 – 70) = 30 Hz, since as per Nyquist theorem maximum input frequency should be less than 100 ÷ 2 = 50 Hz. In general, alias frequency is given by the following equation:

Alias frequency = absolute value of (Closest integer multiple of sampling
frequency – Input frequency)

For example if the maximum input frequency is 170 Hz and sampling frequency is 100 Hz, alias frequency will be (200 – 170) = 30 Hz.

For output frequencies close to half the sampling rate, the output low-pass filter may not adequately remove the sampling image ("alias"), which will appear as a spurious component.

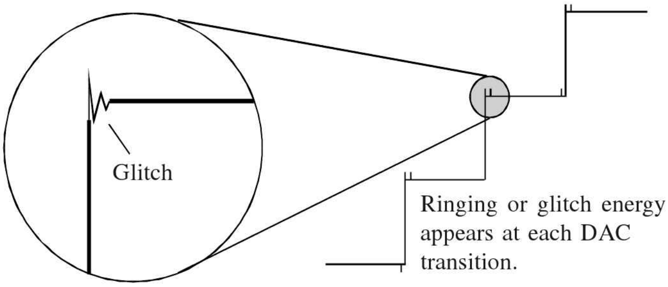
Spurious components. This problem is mainly due to imperfections in DAC, one because of bit weight errors and the other because of transient energy generated (glitches) when DAC changes levels. The later is shown in Figure 7.43. This can be minimized by adding a sample and hold circuit in the output to act as a deglitcher as shown in Figure 7.44.

7.7 DIGITAL SIGNAL GENERATORS

Digital signal generators are broadly divided into the following categories:

- Arbitrary Waveform Generators (AWG/ARB)
- Arbitrary Function Generators (AFG) or Synthesized Function Generators (SFG)
- Data or Pattern Generators (DG)

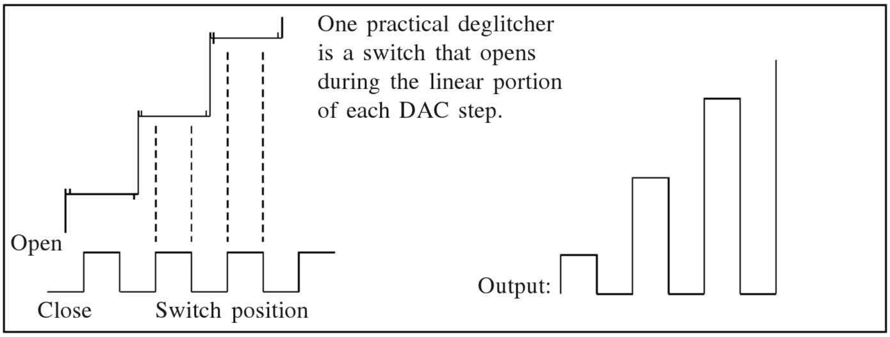
Each of these types has its unique strengths:



After quantization error, the majority of spurious energy from a well-designed DDS occurs due to glitches appearing at the transitions in the digital-to-analog converter.

Ringing or glitch energy appears at each DAC transition.

Figure 7.43 Appearance of glitch in DAC.



One practical deglitcher is a switch that opens during the linear portion of each DAC step.

Figure 7.44 Deglitching technique.

AWG. This can produce any waveform we can imagine. We can use a variety of methods—from mathematical formulae to drawing the waveform to create the needed output.

AFG. Typically this instrument offers waveforms variations, but with excellent stability and fast response to frequency changes. If the UUT requires the classic sine and square waveforms and the ability to switch almost instantly between two frequencies, the AFG is the best instrument. The cost of AFG is less than that of AWG.

DG. This provides long, continuous stream of binary data, with specific information content and timing characteristics.

7.7.1 Arbitrary Waveform Generator

Fundamentally, AWG is a sophisticated *playback* system that delivers waveforms based on stored digital data that describes the constantly changing voltage levels of an AC signal with or without dc content. This data is stored in fast RAM. The data is sampled by actually measuring a waveform with an instrument such as an oscilloscope, or by using graphical or mathematical techniques. The signal can be reconstructed at any time by reading back the memory locations and feeding the data points through a digital-to-analog converter (DAC). Figure 7.45 shows the block diagram of AWG.

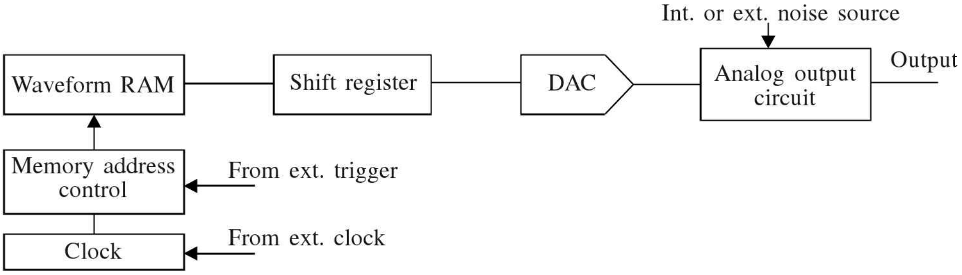


Figure 7.45 Block diagram of AWG.

Nyquist sampling theorem which states that the sampling frequency must be at least twice that of the highest frequency component of the sampled signal, must be satisfied for better fidelity of the signal production. Thus stored data must have enough points to faithfully retrace the details of the desired signal. An AWG can take these points and read them out of memory at any frequency within its specified limits. However, there is a finite maximum frequency, or sample rate, at which instrument can operate. This is usually specified in terms of Megasamples or Gigasamples per second. Today's fastest AWG can achieve about 3 GS/s.

Another way to generate arbitrary waveforms is to use the DDS technique where the RAM includes the amplitude of the desired waveform corresponding to each phase input.

Important parameters of AWGs are explained below.

Vertical resolution (amplitude)

Resolution pertains to the binary word width, in bits, of the instrument's DAC. More bits means higher resolution. While *more is better*, higher-frequency AWGs usually have lower resolution—8 or 10 bits—than that provided by general purpose AWGs offering 12 or 14 bits.

An AWG with 10-bit resolution provides 1024 sample levels spread across the full voltage range of the instrument If, for example, this 10-bit AWG has a total voltage range of $2V_{\text{p-p}}$, each sample represents a step of approximately 2 mV—the smallest increment the instrument can deliver, assuming that it is not constrained by other factors in its architecture.

Memory depth

This plays a key role in the instrument's flexibility. More (deeper) memory provides either of the two benefits:

(i) More cycles of the desired waveform can be stored. This is useful because it reduces the number of *endpoints*. An endpoint is the last memory location occupied by the waveform, after which the AWG must wrap around and return to the beginning in order to continue to produce the output signal. There are unavoidable errors that occur at this transition, so it is desirable to minimize the number of endpoints.

(ii) More waveform details can be stored. Complex waveforms have high frequency information in their pulse edges and transients. It is difficult to interpolate these fast transitions as we did with the simple, predictable sine wave. To faithfully reproduce a complex signal, the available waveform memory capacity must be used to store more transitions and fluctuations rather than more cycles of the signal.

AWGs are available with a few Msample memory depth. When combined with the high sample rate, these instruments can store and reproduce complex RF waveforms, also including pseudo-random bit streams for use in physical layer testing of network equipment. Similarly, these fast AWGs with deep memory can generate extremely short-duration digital pulses and transients.

Frequency and memory depth calculations

The output frequency, f_{out}, is given as

$$f_{\text{out}} = (\text{clock frequency} \div \text{memory depth}) \times (\text{cycles in memory})$$

For example, given a 100 MS/s clock frequency, a memory depth of 4000 samples (points), and assuming that memory contains one waveform cycle, the output frequency will be 25 kHz. But for the same data if memory contains four cycles, the output frequency will be 100 kHz. The number of endpoints transitions will then be reduced. Note that for one stored cycle, the samples are 10 ns apart (horizontal resolution), whereas for four stored cycles, each sample represents 40 ns in time. Therefore, the increase in frequency (and the reduction in endpoint aberrations) comes at the cost of some horizontal resolution.

Example 7.1

An AWG has 12 bit vertical resolution and memory depth of 16,000 points . The number of points in RAM represent one complete cycle of the waveform (or 360°). Plot memory depth versus DAC codes with reference to phase angle.

Solution

Each RAM address corresponds to a phase increment of 360°/points, where points is the waveform length. Sequential RAM addresses will contain the amplitude values for the individual points (0° to 360°) of the waveform. Figure 7.46 shows the plot.

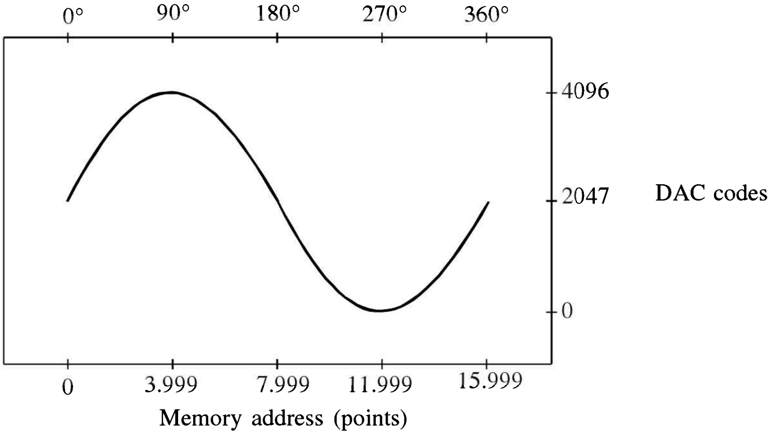


Figure 7.46 Plot between memory addresses and DAC codes.

Modifying the waveform

Once the basic waveform is defined, other operations can be applied to modify or extend it as explained in the following subsections.

Filtering

Filtering allows us to remove selected bands of frequency content from the signal. For example, when testing an analog-to-digital converter (ADC), it is necessary to ensure that the analog input signal (which comes from AWG) is free of frequencies higher than the converter's clock frequency. This prevents unwanted aliasing distortion in the UUT output, which would otherwise compromise the test results.

One reliable way to eliminate these frequencies is to apply a steep low-pass (anti-alias) filter to the waveform. Figure 7.47 shows the placement of filter with DAC.

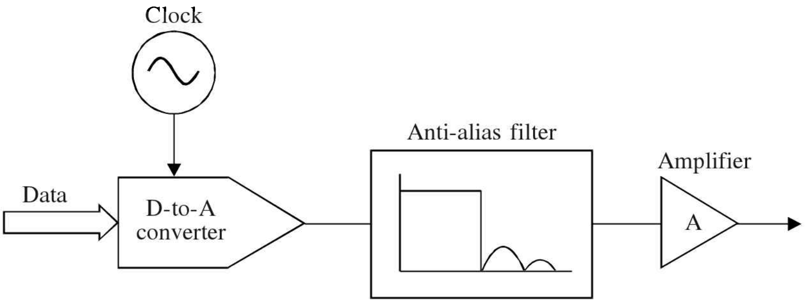


Figure 7.47 Use of anti-alias filter.

For example, for an AWG with 40 MHz fixed sampling frequency, a 17 MHz, ninth-order elliptical filter can provide attenuation in excess of 60 dB for sinusoidal signals greater than 19 MHz. A 10 MHz seventh order Bessel filter can be used for non-sine wave outputs. The Bessel filter provides slower amplitude roll-off for anti-alias filtering, but maintains linear phase response to minimize shape distortion for complex waveshapes. AWG can automatically select the appropriate filter when the output function is selected.

Filters can also be used to *roll-off* pure waveforms such as square and triangular waves. Sometimes it is simpler to modify an existing waveform in this way than to create a new one.

Waveform sequencing

This technique is used to generate different types of waveforms from the already generated basic waves by writing instructions that properly sequence them. So it allows to store a huge number of *virtual* waveform cycles in the instrument's memory. The instructions reside in a separate sequence memory and cause specified segments of the waveform memory to repeat.

For example, let a 4000 point memory be loaded with a clean pulse using 2000 points and a distorted pulse using the balance 2000 points. If we were limited to basic repetition of the memory content, then AWG would always repeat the two pulses, in order, until the command to stop. But waveform sequencing changes all this. Suppose we want the distorted pulse to appear twice in succession 1000 times after every 325 cycles. Figure 7.48 gives the required sequence.

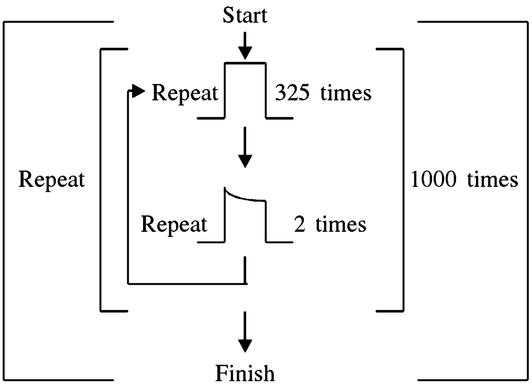


Figure 7.48 Waveform sequencing.

Sequencing provides much improved flexibility without compromising the resolution of the individual waveforms. Note that any sequenced waveform segment must continue from the same amplitude point as the segment preceding it. In other words, if a sine wave segment's last sample value was 1.5 V, the starting value of the next segment in the sequence must be 1.5 V as well. Otherwise, an undesirable glitch can occur when DAC attempts to abruptly change to the new value. This endpoint transient introduces leakage error in the frequency domain because many spectral terms are required to describe the discontinuity. Figure 7.49 illustrates this problem.

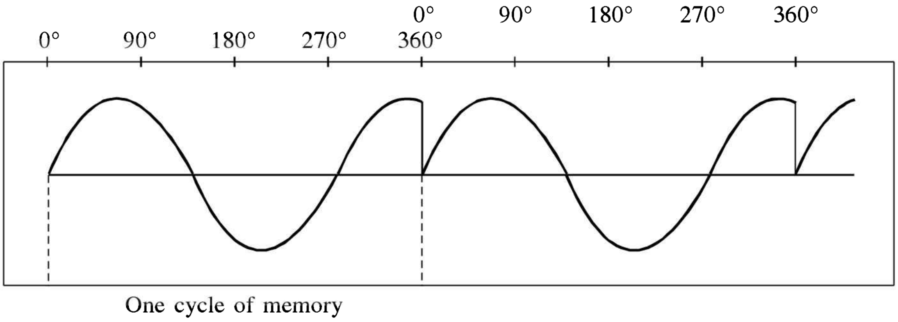


Figure 7.49 Sequencing problem.

Leakage error is caused when the waveform record does not include an integer number of cycles of the fundamental frequency. Power from the fundamental frequency and its harmonics is transferred to spectral components of the rectangular spectral function.

Instead of the expected narrow spectral lines, leakage can cause significant spreading around the desired spectral peaks. We can reduce leakage errors by adjusting the window length to include an integer number of cycles or by including more cycles within the window to reduce endpoint transient size. Some signals are composed of discrete, non-harmonically related frequencies. Since these signals are non-repetitive, all frequency components cannot be harmonically related to the window length. We should be careful in these situations to minimize endpoint discontinuities and spectral leakage.

Tools and methods of creating AWG waveforms

We have discussed the AWG's ability to produce long and complex waveforms. But this would be almost useless if one has to key in point-by-point values for every location in the waveform memory. Fortunately, there are many approaches to waveform creation.

Most lab-quality AWGs are furnished with some standard waveforms, either stored in the instrument's local nonvolatile memory, or provided on disk media. These waveforms make an excellent starting point for modification via the AWG's editing tools.

One of the most efficient ways to derive a new waveform is to *learn* it from a digital storage oscilloscope (DSO). If, for example, stimulus signals are needed for testing a newly developed product in manufacturing, the DSO can be used to capture the waveforms from a known-good prototype. The DSO's acquisition memory contents can then be copied via a GPIB or Ethernet connection between the two instruments. Assuming that the AWG has sufficient bandwidth and resolution, it can duplicate the waveform exactly. The waveform can also be modified, if necessary.

A second waveform creation method is AWG's built-in math tool. All "standard" waveforms (sine, square, triangle etc.) can be derived from relatively simple mathematical formulae. Some AWGs have math editor that apply convolution, integration and other operations. Many AWGs accept data from PC-based waveform simulation tools.

The third method is by using a built-in graphical editor. Some AWGs offer a means of actually drawing (on the instrument's own display screen) the waveform features we need. Of course, there are finite limitations—we still can not make an edge transition in 0 ps—but the instrument will reproduce all of the waveform details within its performance range.

Some AWGs can actually deliver an analog waveform output and a correlated digital signal at the same time. These instruments have *marker* outputs that can provide either a trigger pulse or a programmed binary expression when a specific sample value is clocked out of the memory.

7.7.2 Arbitrary Function Generator

The Arbitrary Function Generator (AFG) shares many features with the AWG, although it is by design, a more specialized instrument with a narrower range of applications. The AFG offers unique strengths: it produces stable waveforms in standard shapes—particularly all-important sine and square waves that are both accurate and agile. Agility is the ability to change quickly and cleanly from one frequency to another. Most AFGs offer sine, square, triangle, sweep, pulse, ramp, modulation, and noise. AFGs are designed to provide improved phase, frequency and amplitude control of the output signal. Many can modulate the signal from internal or external sources, which is essential for some type of standards compliance testing.

In the past, AFGs created their output signals using analog oscillators and signal conditioning. More recent AFGs rely on a digital architecture and have much in common with the AWGs. The same basic building blocks are there: waveform memory, DAC, and output signal conditioning. And like the best AWGs, the AFG uses a Direct Digital Synthesis (DDS), discussed above, to determine the rate at which samples are clocked out of the memory.

7.7.3 Data Generator

Unlike AWG and AFG, which are used to generate waveforms with analog shapes and characteristics, DG generates volumes of binary information. Also known as pattern generator, the DG produces the stream of 1s and 0s needed for testing computer buses, microprocessors and other digital systems. Some of the important features of DG are:

Sequencing. An absolute necessity in the world of data and pattern generation. No internal memory can be deep enough to store millions of pattern words (also known as *vectors*) required for a thorough digital device test. Consequently, DGs are equipped with sophisticated sequencers, far more than those of other type of sources.

Multiple outputs. Where the AWG or AFG may have two or four outputs, the DG may have hundreds of output channels to support numerous data, address and control lines of a typical digital device/system.

Pattern data sources. The DG must accept data from logic analyzers, DSOs, simulators and even spreadsheets. This is because a complex digital pattern would be impossibly tedious and error prone if entered by hand. Moreover digital data is usually available from various simulation and verification steps in the design process.

Display. The display of DG must emphasize the details of many channels of pattern data simultaneously rather than the signal amplitude vs. time. It should offer markers, scrolling and other time saving features to help the user focus on the data of interest.

Figure 7.50 shows the block diagram of a DG.

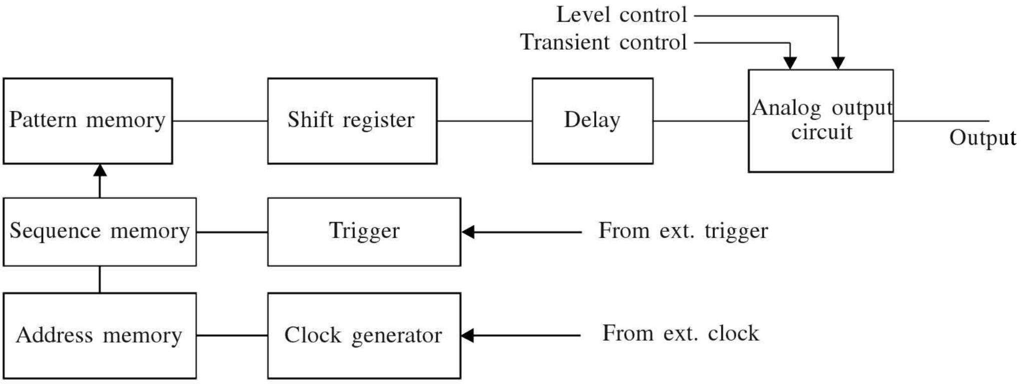


Figure 7.50 Block diagram of DG.

It can be noted from Figure 7.50 that DAC is missing. But, in fact, DAC is not required because the DG does not need to trace out the constantly shifting levels of an analog waveform. Analog output circuit is used to set voltage and edge parameters that apply to the whole pattern. For example, most DGs provide a way to program the logic “1” and “0” voltage levels for the pattern and also rise time of the pulse. These are used to simulate erratic voltage levels and slow edge rise times.

The delay circuit is responsible for implementing the small changes in edge positioning that support both jitter (explained later) and timing tests. It can deliver tiny changes (of the order of picoseconds) in edge placement. For example, some DGs provide simple front panel controls that allow to move all edges or selected edges in 5 ps steps within a range of ±100 ps. These small timing changes model the classic jitter phenomenon in which the placement in time of a pulse edge moves erratically about a nominal centre point. One can test jitter tolerance by changing and observing the effects of edge timing in relation to the clock.

The delay circuit is also used in testing timing problems such as setup-and-hold violations. Most clocked devices require the data signal to be present for a few nanoseconds before the clock pulse appears (setup time) and to remain valid for a few nanoseconds (hold time) after the clock edge. The delay circuit makes it easy to implement this set of conditions. Just as it can move a signal edge a few picoseconds at a time, it can move that edge in hundreds of picoseconds, or in nanoseconds. This is what is needed to evaluate setup and hold time.

The sequencer of DG must provide tremendously long and complex patterns and must respond to external events—usually a UUT output condition that prompts a branch execution in the DG sequencer. DG’s pattern memory capacity, typically about 256 Kbits, may seem modest in comparison to the largest AWG waveform memory. But like the AWG sequencer, the DG sequencer can loop on short pattern segments to produce a data stream of much greater length. It can wait for an external event or trigger, then execute a series of repeat counts or conditional jumps. In addition, the DG sequencer offers many levels of loop nesting and branch conditions. It is a tool that can be controlled using normal programming conventions, and produces address, data, clock and control signals for almost any digital device. Figure 7.51 shows how a few short instructions and pattern segments can unfold into millions of lines of stimulus data.

7.8 KEY CHARACTERISTICS OF DIGITAL SIGNAL GENERATORS

Sample rate

This is the maximum clock or sample rate at which the instrument can operate. The sample rate controls the frequency and fidelity of the main output signal. Nyquist theorem states that the clock rate must be at least twice the frequency of the desired signal’s highest component frequency to ensure accurate reproduction. Four, eight or even hundred times oversampling may be necessary in some applications.

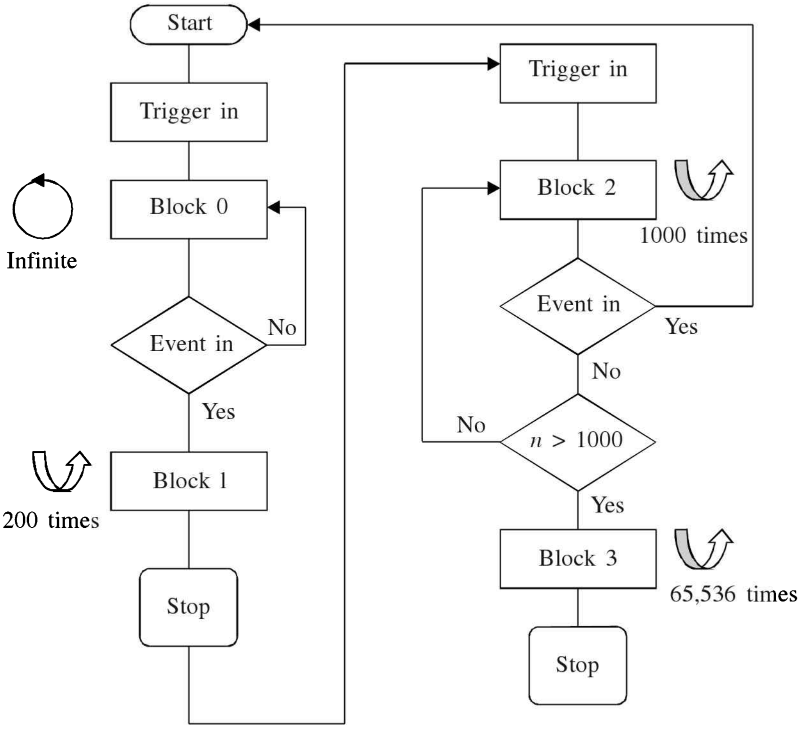


Figure 7.51 DG sequence flow, showing loops and conditional events.

Vertical resolution

The vertical resolution of the DAC defines the amplitude accuracy and distortion of the reproduced waveform. A DAC with inadequate resolution contributes to quantization errors, causing imperfect waveform generation.

Memory depth

This limits the length of a non-repeating waveform. It contributes to signal fidelity because using a low frequency signal (relative to the sample clock rate) permits more data points to be stored per cycle. Without sufficient memory depth, we cannot take advantage of the high relative sample rate. Many AWGs offer 2, 4, even 8 Msamples of memory.

Number of channels

Many applications require more than one output channel. AWGs with up to 4 independent channels are available. This feature is indispensable for simulating sensors in automotive (anti-lock brakes) or bio-electrical applications, and more. Some AWGs have one or two analog outputs and up to 14 high-speed digital outputs for mixed-signal testing.

In addition to these key specifications, some of the other features that can make the instrument easier to use, more productive or more cost effective are as follows:

Modifying AWG waveforms in real time: integrated editors. Suppose we need a series of waveform segments having the same shape but different amplitudes as the series proceeds. We can recalculate the waveform or redraw it using an off-line waveform editor. But both these approaches are unnecessarily time consuming and error prone. A better method is to use an integrated editing tool that can modify the waveform memory in both the time and amplitude domains. These editors automatically update the sample memory to deliver near real-time changes in the output signal.

Creating application specific waveforms: dedicated functions. AWGs are ideal for generating complex test waveforms. Some AWGs provide application specific signal creation functions. Predictable jitter simulation is an application that cuts across many industries. The ability to inject jitter into the waveform file can be a time consuming, math-intensive effort. Some of the AWGs offer special "jitter editors" to assist in adding jitter to any appropriate signal. These editors make a complex process simple: define the digital data stream, select the jitter profile and enter the jitter deviation and frequency. The instrument calculates the changes and saves them to the waveform file.

Expanding the waveform length: sequencing. Often, it is necessary to create long waveform files to fully exercise the UUT. Where portions of the waveform are repeated, a waveform sequencing function can save a lot of tedious, memory intensive waveform programming. Sequencing allows us to fill in a table listing the waveforms in order of appearance. Programmable repeat counters, branching on external events and other control mechanisms determine the number of operational cycles and the order in which they occur. With a sequence controller, we can generate waveforms of almost unlimited length.

Synchronizing analog and digital signals: markers. Marker outputs provide a binary signal that is synchronous with the main analog output signal of an AWG. In general, markers allow us to output a pulse (or pulses) synchronized with a specific waveform memory location. Marker pulses are usually used to synchronize the digital portions of a UUT while the unit is simultaneously stimulated by the analog signal.

Using waveforms from other resources: importing data. Data import function allows to use waveform files created outside the AWG. For example, a waveform captured by a digital storage oscilloscope can be easily transferred via GPIB or Ethernet to the AWG. Simulators and other tools are also available as sources of waveforms.

QUESTIONS

7.1 Explain the principle of oscillations of basic sine wave generators.

7.2 What is meant by frequency stability, phase stability and amplitude stability of oscillators? What are the causes of these instabilities and how are they taken care of?

7.3 A Wien bridge oscillator of Figure 7.4 has R = 10 kΩ, C = 10 nF and R_1 = 1 kΩ. What is the minimum value of R_2 for oscillations to start and what is the frequency of oscillations? What are the techniques used to stabilize the amplitude of oscillations?

7.4 Refer to Figure 7.5 and design an *RC* phase shift oscillator for a frequency of 100 kHz. Let the value of C be 100 pF. Choose appropriate values of resistors R_1 and R_2.

7.5 Explain the general configuration of radio frequency signal generators and derive conditions for oscillations.

7.6 Refer to Figure 7.7 and design a Hartley oscillator for a frequency of 100 kHz. Given a 50 mH centre-tapped coil, choose appropriate values of resistors R_1 and R_2.

7.7 Refer to Figure 7.8 and design a Colpitts oscillator for a frequency of 100 kHz. Given a 50 mH coil, choose the appropriate values of resistors R_1 and R_2.

7.8 A crystal is specified to have L = 22 mH, C_s = 0.018 pF, C_p = 4.5 pF and R = 30 Ω. Find f_s, f_a, and Q.

7.9 Refer to Figure 7.11 and design a crystal oscillator using the crystal as given in Question 7.8. Choose appropriate values of resistors R_1 and R_2 and capacitors C_1 and C_2.

7.10 Differentiate between bistable, astable, and monostable multivibrators.

7.11 Explain the Schmitt trigger circuit.

7.12 Refer to Figure 7.13. If the opamp has saturation voltages of ± 10 V and the value of R_1 = 10 kΩ, what is the value of R_2 so that the threshold voltages of the bistable multivibrator are ±5 V?

7.13 Explain the operation of the circuit of Figure 7.13 if an inverting input is connected to the ground and input v_i is applied through R_1.

7.14 Refer to Figure 7.17(a). If the opamp has saturation voltages of ± 10 V and the values of R_1 = 100 kΩ and $R_2 = R$ = 1 MΩ, what is the value of C for the frequency of oscillation to be 274 Hz? Draw the output waveform.

7.15 Refer to Figure 7.19(c). If the opamp has saturation voltages of ± 10 V and the values of R_1 = 10 kΩ and C = 0.01 μF, find the values of R and R_2 so that the frequency of oscillation is 1 kHz and the triangular waveform has 5 V peak-to-peak amplitude. Draw the output waveform.

7.16 Refer to Figure 7.21(a). If the opamp has saturation voltages of ± 10 V and the values of C = 0.1 μ*F*, β = 0.1, and V_{D1} = 0.7 V, find the value of R that results in a 100 μs output pulse.

7.17 Draw the block diagram of a function generator which generates pulse, triangular and sinusoidal waveforms.

7.18 Define the various pulse parameters used in the specifications of signal generators.

7.19 What is jitter? Define the various types of jitters.

7.20 What are various synthesis techniques which generate frequencies from a given reference frequency? List the figure of merits used to choose a synthesizer.

7.21 Draw the block diagram and explain the operation of a Direct Analog Synthesizer to generate output between 1 and 59 MHz, given a clock of 10 MHz.

7.22 Draw the block diagram and explain the operation of a fast switching Direct Analog Synthesizer to generate output between 1 and 27 MHz, given a clock of 1 MHz.

7.23 Design a Programmable Decade Indirect Synthesizer to synthesize an output of 8.46 MHz from a reference source of 2 MHz.

7.24 Explain the operation of a Standard Indirect Frequency Synthesizer using a swallow counter.

7.25 Draw the block diagram of a Direct Digital Synthesizer. Explain its operation. What are its advantages and limitations over other techniques?

7.26 Draw the block diagram of an Arbitrary Waveform Generator (AWG). Explain its operation. What is meant by vertical resolution?

7.27 Why is memory depth important in selecting an Arbitrary Waveform Generator (AWG)? How are frequency and memory depth calculated?

7.28 An AWG has 10 bit vertical resolution and memory depth of 10,000 points. The number of points in RAM represent one complete cycle of the waveform (or 360°). Plot memory depth versus DAC codes with reference to phase angle.

7.29 What are the tools and methods used to create and modify waveforms in an AWG?

7.30 Draw the block diagram of a digital generator. Explain its operation. How is it different from AWG and AFG? What are its important features?

7.31 What are the key characteristics of digital signal generators?

Chapter 8

Signal Analyzers

8.1 INTRODUCTION

A signal may consist of a dc, or a pure sinusoidal component, or a waveform consisting of many frequency components. For example:

- A dc power supply is supposed to have a pure dc output. But it always contains some distortion in the form of harmonics of the input frequency.
- An amplifier operation should be linear. But due to the nonlinear behaviour of devices and components, the output is distorted.
- A modulated signal consists of a number of frequencies.

Instruments are available to measure distortion in the signal and to represent the energy of each frequency component of the signal.

A signal can be analyzed in the time domain, for example, by using a CRO, or in the frequency domain, for example, by using a spectrum analyzer. Time domain analysis has the disadvantage that we cannot see the contribution of individual frequency component of a complex waveform. Signal characteristics that are difficult to see in the time domain frequently become visible in the frequency domain. A classic example is the square wave, which appears as the sum of many harmonics in the frequency domain. When we want to see the spectrum content of the signals, there are several choices of the test equipment—spectrum analyzers, Fast Fourier Transform (FFT) analyzers, vector analyzers, and digitizing oscilloscopes (DSO). Each has its own advantages and disadvantages and the choice depends upon the application.

Further, for design and analysis of digital systems, a logic analyzer is an important instrument. This instrument provides visualization of what is happening inside the digital devices or digital systems. For example, analyzing the activities on the microprocessor data/address busses and control and I/O lines.

In this chapter, we will introduce concepts related to distortion and time versus frequency representation and then study the construction and features of some of the signal analyzers.

8.2 DISTORTION

Distortion is a measure of signal impurity. It is usually expressed as a percentage or decibel ratio of the undesired components to the desired components of a signal. Distortion of a device is measured by feeding into it one or more sine waves of various amplitudes and frequencies. In simple terms, frequencies in the output which were not present in the input constitute distortion.

However, strictly speaking, components due to power line interference or other spurious signals do not constitute distortion. There are many methods in common use of measuring distortion: harmonic distortion and intermodulation distortion. These are different test procedures rather than different forms of distortion in the device under test.

8.2.1 Harmonic Distortion

Harmonic distortion results when a voltage at a single frequency f, applied to a nonlinear device or system, creates spurious voltages at frequencies (harmonics) $2f$, $3f$, ... , Nf. Of most concern, because they are the closest to the desired signal, are the second and third harmonics. The order of the distortion product is given by the frequency multiplier, for example, the second harmonic is a second-order product. Individual components of harmonics distortion are defined as follows:

$$\%\text{ 2nd harmonic distortion} = \frac{|B_2|}{|B_1|} \times 100 = D_2$$

$$\%\text{ 3rd harmonic distortion} = \frac{|B_3|}{|B_1|} \times 100 = D_3$$

$$\%\text{ nth harmonic distortion} = \frac{|B_N|}{|B_1|} \times 100 = D_N$$

where B_1 is the rms value of the fundamental component, B_2, B_3, ..., B_N are the rms values of the corresponding harmonics.

Total Harmonic Distortion (THD), D, is defined as

$$\%D = \frac{\sqrt{B_2^2 + B_3^2 + \cdots + B_N^2}}{\sqrt{B_1^2}} \times 100 = \sqrt{D_2^2 + D_3^2 + \cdots + D_N^2}$$

This equation is practically inconvenient. Since it is difficult to find the amplitude of the fundamental alone in the presence of harmonics, working equation of THD is defined as

$$\% \text{ THD} = \frac{\sqrt{B_2^2 + B_3^2 + \cdots + B_N^2}}{\sqrt{B_1^2 + B_2^2 + B_3^2 + \cdots + B_N^2}} \times 100$$

The denominator is the amplitude of the signal with all harmonic components.

Further, to take care of the noise component in the signal, we define THD with noise as follows:

THD + n

It is defined as

$$\% \text{ THD} = \frac{\sqrt{B_2^2 + B_3^2 + \cdots + B_N^2 + n^2}}{\sqrt{B_1^2 + B_2^2 + B_3^2 + \cdots + B_N^2 + n^2}} \times 100$$

where n is the amplitude of the noise component.

Signal + noise + distortion (SINAD)

SINAD is a parameter which provides a quantitative measurement of the quality of an audio signal from a communication device. It is the ratio of the total signal power level (desired signal + noise + distortion or SND) to unwanted signal power (noise + distortion or ND). The higher this ratio, the better the quality of the audio signal. The ratio is expressed in dB as 10 log (SND/ND). Remember that this is a power ratio, not a voltage ratio.

8.2.2 Intermodulation Distortion (IMD)

This occurs when the nonlinearity of a device or system with multiple input frequencies causes undesired outputs at other frequencies. In a communication system, this means that signals in one channel can cause interference with adjacent channels. As the spectrum becomes busier and the channels become more tightly spaced, minimizing intermodulation distortion becomes more important.

Given two input signals at frequencies f_1 and f_2, we can compute several significant intermodulation distortion products from the equation

$$Mf_1 \pm Nf_2$$

where $M, N = 1, 2, 3, \ldots$.

The order of the distortion product is given by the sum $M + N$. Figure 8.1 lists some of the harmonics and intermodulation components of a system.

Of the infinite number of distortion products described by this equation, one is of special significance, i.e. the third order products

$$2f_1 - f_2 \quad \text{and} \quad 2f_2 - f_1$$

These products are important because they exist on either side of the original signals f_1 and f_2 and they cannot be removed by filtering in a narrow band system.

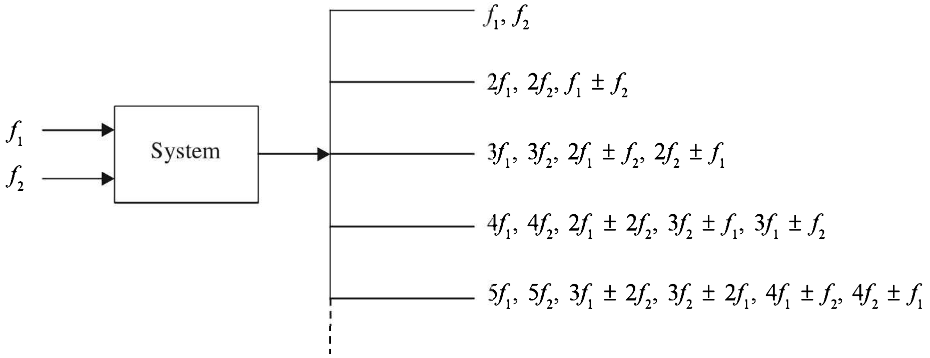

f_1
f_2
System
f_1, f_2
$2f_1, 2f_2, f_1 \pm f_2$
$3f_1, 3f_2, 2f_1 \pm f_2, 2f_2 \pm f_1$
$4f_1, 4f_2, 2f_1 \pm 2f_2, 3f_2 \pm f_1, 3f_1 \pm f_2$
$5f_1, 5f_2, 3f_1 \pm 2f_2, 3f_2 \pm 2f_1, 4f_1 \pm f_2, 4f_2 \pm f_1$

Figure 8.1 Some of the harmonics and IMD components.

Signals transmitted with excessive third-order IMD can interfere with other transmissions. Receivers must also be distortion-free, especially in the preamplifier stages, to prevent crosstalk between adjacent channels.

For example, let the transmitter and receiver frequencies of a cellular phone be as follows:

$$\text{Tx} = 824\text{ MHz} - 849\text{ MHz}, \qquad \text{Rx} = 869\text{ MHz} - 894\text{ MHz}$$

For third-order IMD, assume two transmitter frequencies as

$$f_1 = 824\text{ MHz} \quad \text{and} \quad f_2 = 849\text{ MHz}$$

Then $2f_1 - f_2 = 799$ MHz and $2f_2 - f_1 = 874$ MHz.

This unwanted frequency (874 MHz) lies within the receiver band and will be picked by the phone.

8.3 DISTORTION MEASURING INSTRUMENTS

We will study some of the instruments to measure THD, IMD and spectrum of a signal.

8.3.1 Distortion Analyzer

This is the instrument used to measure total harmonic components. It removes the fundamental component of the signal under investigation and measures the remainder. A block diagram of such an instrument is shown in Figure 8.2. With the switch S_1 at ‘Set Level’, the band-pass filter is adjusted to the fundamental frequency and attenuator is adjusted to obtain full-scale reading on the meter. The switch position is then changed to ‘Distortion’, the notch filter is tuned to the fundamental and attenuator is adjusted again to get full-scale reading on the meter. THD is then calculated from the attenuator position. SINAD can also be measured with this equipment.

Any practical signal contains some hum and noise, and the distortion analyzer will include these in the reading. Because of these added components, the correct term for this measurement is total harmonic distortion and noise (THD + n).

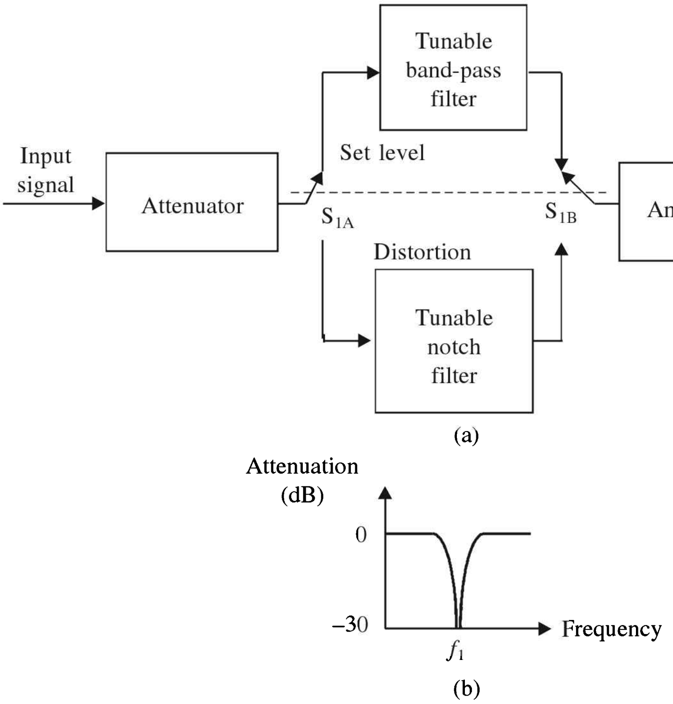


Figure 8.2 (a) Distortion analyzer and (b) typical characteristics of a notch filter.

Additional filters are included on most distortion analyzers to reduce unwanted hum and noise. These usually consist of one or more high-pass filters (400 Hz is almost universal) and several low-pass filters. Common low-pass filter frequencies are 22.4 kHz, 30 kHz, and 80 kHz.

The use of a sine wave test signal and a notch-type distortion analyzer has the distinct advantage of simplicity in both instrumentation and use. This simplicity has an additional benefit in ease of interpretation.

The shape of the output waveform from a notch-type analyzer indicates the slope of the nonlinearity. Displaying the residual components on the vertical axis of an oscilloscope and the input signal on the horizontal gives a plot of the transfer characteristic's deviation from a best-fit straight line. This technique is shown in Figure 8.3. The trace will be a horizontal line

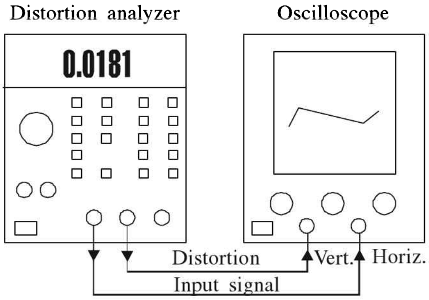


Figure 8.3 Observing the transfer characteristics.

for a perfectly linear device. If the transfer characteristic curves upwards on positive input voltages, the trace will bend upwards at the right-hand side. Examination of the distortion components in real time on an oscilloscope will show such characteristics as oscillation on the peaks of the signal, crossover distortion, and clipping. This is a valuable tool in the design and development of audio circuits and one which no other distortion measurement method can fully match.

8.3.2 Wave Analyzer

This instrument is used to measure each harmonic component individually. Figure 8.4 shows the block diagram of wave analyzer.

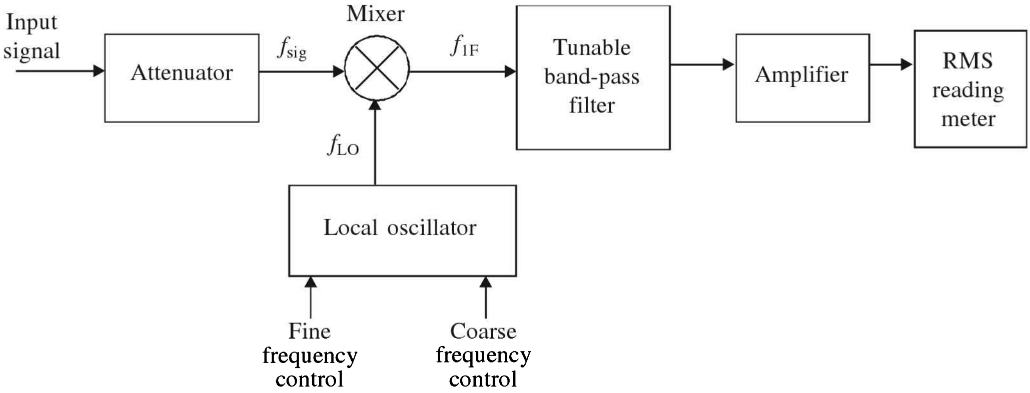


Figure 8.4 Wave analyzer.

This instrument uses the principle of heterodyne (heterodyne means to mix) where signal of frequency, f_{sig}, is mixed with a local oscillator frequency, f_{LO}, generating frequencies $f_{LO} + f_{sig}$, $f_{LO} - f_{sig}$ and harmonics. A narrow band-pass filter is tuned to pass signal of frequency $f_{LO} + f_{sig}$ to the amplifier which is then read by the meter. By varying the frequency of the local oscillator, signal amplitudes of fundamental, 2nd harmonic, 3rd harmonic, etc. can be measured individually. Note that each harmonic frequency is converted to a constant frequency, which allows the use of highly selective filters.

Since the frequency components are measured one at a time, this is called non-real time analyzer (NRTA).

The bandwidth of heterodyne wave analyzer is usually constant. This makes the analysis very difficult if signal frequency does not remain constant during the time required by the instrument for complete analysis. For example, let the bandwidth of the band-pass filter be 10 Hz. If the fundamental frequency of 1 kHz changes by 2%, the 5th harmonic frequency will change by 100 Hz, which is well outside the bandwidth of the band-pass filter.

8.3.3 Intermodulation Analyzer

The commonly accepted solution to the limitations of THD analysis is to measure distortion by the interaction or intermodulation (IM) of two or more signals passing through a device or system simultaneously. Many methods have been devised to measure this interaction.

SMPTE IM analyzer

This is the most common method to measure IM. It is named after the Society of Motion Picture and Television Engineers (SMPTE), which first standardized its use. IM measurements according to the SMPTE method have been in use since the 1930s. The test signal is a low-frequency tone (usually 50 Hz) and a high-frequency tone (usually 7 kHz) mixed in a 4:1 amplitude ratio. Other amplitude ratios and frequencies are used occasionally. The signal is applied to the device or system under test, and the output signal is examined for modulation of the upper frequency by the low-frequency tone. The amount by which the low-frequency tone modulates the high-frequency tone indicates the degree of nonlinearity. As with harmonic-distortion measurement, this may be done with a dedicated distortion analyzer. The modulation components of the upper signal appear as sidebands spaced at multiples of the lower-frequency tone as illustrated in Figure 8.5. The amplitudes of the sidebands are rms-summed and expressed as a percentage of the upper-frequency level.

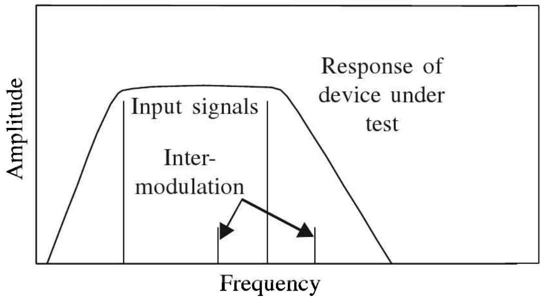


Figure 8.5 Intermodulation in band-limited signal.

Figure 8.6 shows the block diagram of an SMPTE IM analyzer. A dedicated distortion analyzer for SMPTE testing is quite straightforward. The signal to be analyzed is high-pass filtered to remove the low frequency tone. The sidebands are demodulated using an amplitude modulation detector. The result is low-pass filtered to remove the residual carrier components. Since this low-pass filter restricts the measurement bandwidth, noise has little effect on SMPTE measurements. The analyzer is very tolerant of the harmonics of the two input signals, allowing fairly simple oscillators to be used. It is important that none of the harmonics of the low-frequency oscillator occur near the upper-frequency tone, since the analyzer will view these as distortion. After the first stage of high-pass filtering in the analyzer there is little low-frequency information left to create intermodulation in the analyzer itself. This simplifies the design of the remaining circuitry.

CCIT (DFD) IM analyzer

Twin-tone intermodulation or International Telephone Consultative Committee (CCIT) or Difference Frequency Distortion (DFD) is another method of measuring distortion by using

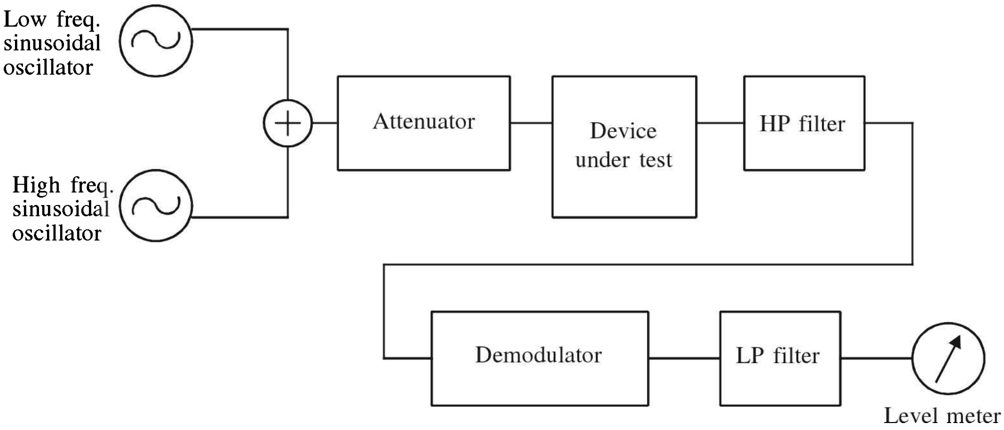


Figure 8.6 Block diagram of SMPTE IM analyzer.

two sine waves. The test signal consists of two closely spaced high-frequency tones as shown in Figure 8.7. When these are passed through a nonlinear device, IM products are generated at frequencies related to the difference in frequency between the original tones, i.e. $f_1 - f_2$, $2(f_1 - f_2)$, $3(f_1 - f_2)$, etc. and $2f_1 - f_2$, $2f_2 - f_1$, $3f_1 - 2f_2$, $3f_2 - 2f_1$, etc. For the typical case of signals at 14 kHz and 15 kHz, the IM components will be at 1 kHz, 2 kHz, 3 kHz, etc. and 13 kHz, 16 kHz, 12 kHz, 17 kHz, and so on. Even-order, or asymmetrical distortions produce low frequency difference-frequency components. Odd-order, or symmetrical, nonlinearities produce components near the input-signal frequencies. The most common application of this test measures only the even-order components since they may be measured with only a multipole low-pass filter. Measurement of the odd-order components requires a spectrum analyzer or a selective voltmeter. The measurement residuals in the two cases are approximately 110 dB and 85 dB below the input signals, respectively. The CCIT test has several advantages over either harmonic or SMPTE IM testing. The signals and distortion components may almost always be arranged to be in the pass-band of the device under test. At low frequencies, the required spacing becomes proportionally smaller, requiring a higher resolution in the spectrum analysis. At such frequencies, a THD measurement may be more convenient.

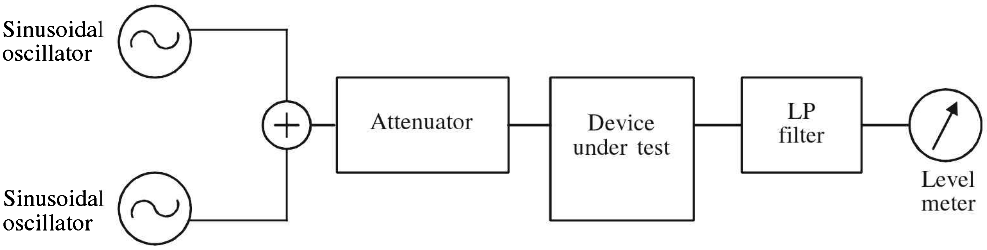


Figure 8.7 CCIT (DFD) IM analyzer.

8.4 SPECTRUM ANALYZER

An analysis of a complex waveform, prepared in terms of a graphic plot of the amplitude versus frequency, is known as *spectrum analysis*. Spectrum analysis recognizes the fact that waveforms are composed of the summation of a group of sinusoidal waves, each of an exact frequency and all existing together simultaneously. Three axes (amplitude, time, and frequency) are important when considering varying frequency. The time-domain (amplitude versus time) plot is used to consider phase relationships and basic timing of the signal and is normally observed with an oscilloscope. The time domain is better for many measurements, and some measurements can be made only in the time domain.

For example, pure time-domain measurements include pulse rise and fall times, overshoot, and ringing. The frequency-domain (amplitude versus frequency) plot is used to observe frequency response—a spectrum analyzer is used for this purpose.

Figure 8.8 illustrates the differences between frequency- and time-domain plots of a fundamental frequency, f_1, and its second harmonic, $2f_1$, with respect to time, frequency, and amplitude. The time-domain display is as it would be seen on an oscilloscope. The solid (dark) curve, $f_1 + 2f_1$, is the actual display. The solid (light) curve, f_1, and the dashed curve, $2f_1$, are drawn to illustrate the fundamental and second harmonic frequency relationship used to formulate the composite signal $f_1 + 2f_1$. The frequency-domain display is as it would be seen on a spectrum analyzer. Note that the components of the composite signal are clearly seen.

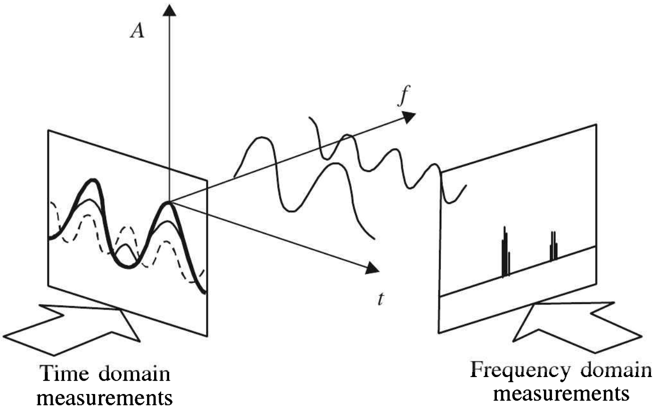
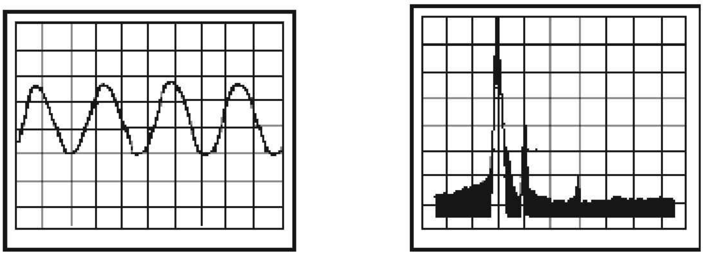


Figure 8.8 Relationship between time and frequency domains.

8.4.1 Frequency-Domain Display Capabilities

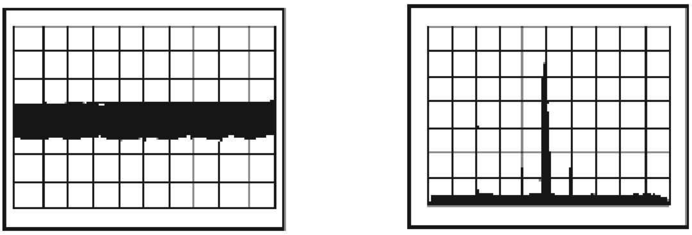
The frequency domain contains information not found in the time domain. The spectrum analyzer can display signals composed of more than one frequency (complex signals). It can also discriminate between the components of the signal, and measure the power level at each one. It is more sensitive to low-level distortion than an oscilloscope. Its sensitivity and wide dynamic range are also useful for measuring low-level modulation, as illustrated in Figures 8.9(a) and 8.9(b). The spectrum analyzer is useful in the measurement of long- and short-term stability such as noise sidebands of an oscillator, residual FM of a signal generator, or frequency drift of a device during warm-up.



Oscilloscope display

Spectrum analyzer display

(a) Low-level distortion



Oscilloscope display

Spectrum analyzer display

(b) Low-level modulation

Figure 8.9 Examples of time-domain (left) and frequency-domain (right) low-level signals.

The modern spectrum analyzers have analysis capabilities from a few Hz to well over 100 GHz in frequency and an amplitude range of well over 100 dB.

Since these analyzers display the complete spectrum of the signal frequencies simultaneously on the same screen, they are called real time analyzers (RTA) or real-time spectrum analyzers (RTSA).

8.4.2 Types of Spectrum Analyzers

Real-time spectrum analyzers are realized in basic two types:

- Filter bank spectrum analyzer
- Superheterodyne spectrum analyzer

Filter bank spectrum analyzer

Figure 8.10 shows the block diagram of this type of analyzer. It has a bank of fixed tuned filters spread over the frequency range of interest. Each filter passes one frequency. A detector then measures the power level of the signal passing through the filter, producing a dc voltage that drives the vertical plates of the display. An electronic sweep switch selects one output of the detector at a time. As the ramp (sweep) generator sweeps through its frequency range, a trace is drawn across the screen. This trace shows the spectral content of the input signal within a selected range of frequencies.

Depending on the desired frequency range and resolution, this analyzer may require large quantity of hardware.

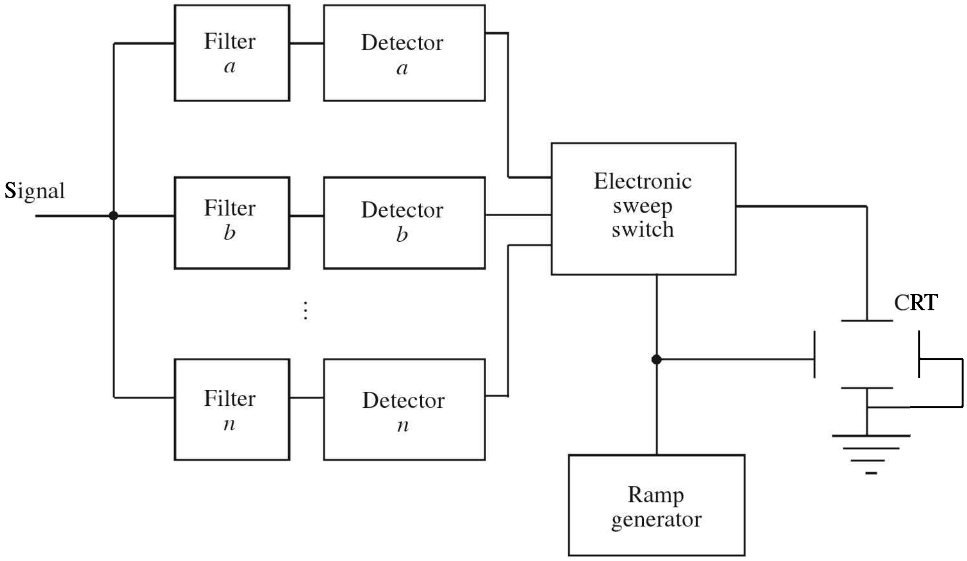


Figure 8.10 Filter bank spectrum analyzer.

Superheterodyne spectrum analyzer

Superheterodyne means to mix the frequencies in the super (above) audio range.

Figure 8.11 is a simplified block diagram of a superheterodyne spectrum analyzer. As shown in the figure, the input signal is first attenuated, if required. Then it passes through a low-pass (LP) filter. Then it is mixed with a signal from the local oscillator (LO). Because the

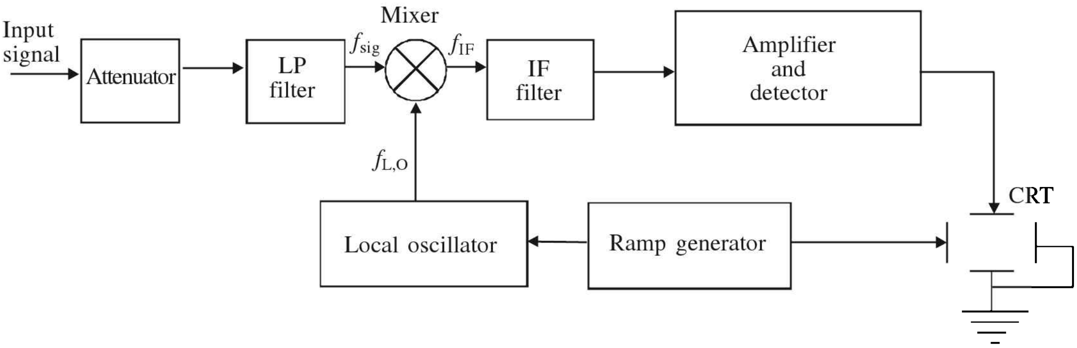


Figure 8.11 Simplified block diagram of a spectrum analyzer.

mixer is a non-linear device, its output includes not only the two original signals but also their harmonics and the sums and differences of the original frequencies and their harmonics. The mixed signal falling within the pass-band of the intermediate-frequency (IF) filter, are amplified (sometimes log amplified to increase the range). The resultant signal is then applied to the vertical plates of a cathode-ray tube (CRT) after it is rectified by the envelope detector. It produces a vertical deflection on the CRT screen (the display). A ramp generator deflects the CRT beam horizontally across the screen from left to right. The ramp also tunes the local oscillator, so that its frequency changes in proportion to the ramp voltage. So we get an X-Y display on the screen. The vertical axis is calibrated in amplitude. Virtually all analyzers offer the choice of a linear scale calibrated in volts or a logarithmic scale calibrated in dB (some analyzers also offer a linear scale calibrated in units of power.) The log scale is used far more often than the linear scale because the log scale has a much wider usable range. The log scale allows signals as far apart in amplitude as 70 to 100 dB to be displayed simultaneously. On the other hand, the linear scale is usable for signals differing by no more than 20 to 30 dB. We can measure either the absolute value of a signal or the amplitude difference between any two signals.

Since the ramp generator controls both the horizontal position of the trace on the display and the LO frequency, we can calibrate the horizontal axis of the display in terms of input-signal frequency.

Selection of frequencies

Let us see how to pick up the LO frequency and the IF to design an analyzer with the desired frequency range? Let the tuning range be 0 to 2.9 GHz. If we choose IF to be within the input signal range, say 1 GHz, then since this frequency is within our desired tuning range and since the output of the mixer also includes the original input signal, an input signal at 1 GHz would give us a constant output from the mixer at the IF. The 1 GHz signal would thus pass through the system and give us a constant vertical deflection on the display regardless of the tuning of the LO. The result would be a hole in the frequency range at which we could not properly examine signals because the display deflection would be independent of the LO. So we should choose IF above the highest frequency to which we wish to tune. Let this frequency be 3.6 GHz.

The IF filter is either tuned to the sum or the difference of LO and input signal frequencies. Let us assume that the IF filter is tuned to difference of the LO and signal frequency. Then, the tuning equation of the analyzer is given by

$$f_{IF} = f_{LO} - f_{sig} \tag{8.1}$$

where f_{sig} is the signal frequency, f_{LO} is the local oscillator frequency, and f_{IF} is the intermediate frequency (IF).

The LO frequency, from Eq. 8.1, will be $f_{IF} + f_{sig}$. So, once IF is decided and the range of input frequency is known, the LO tuning frequency range can be calculated. In our example for input signal range from 0 to 2.9 GHz and IF of 3.6 GHz, the LO frequency range is 3.6 GHz to 6.5 GHz.

Why low-pass filter?

Let us assume that the signal frequency exceeds the range of LO frequency. Equation (8.1) can then be written as

$$f_{\mathrm{IF}} = f_{\mathrm{sig}} - f_{\mathrm{LO}} \tag{8.2}$$

As per Eq. (8.2), the input frequency variation from 7.2 GHz to 10.1 GHz (i.e. $f_{\mathrm{sig}} + 2f_{\mathrm{IF}}$) and LO frequency from 3.6 to 6.5 GHz again results in IF of 3.6 GHz. That is, the architecture which is designed for input frequency range of 0 to 2.9 GHz, also works for tuning range of 7.2 GHz to 10.1 GHz. This frequency range (called *image frequency*) is filtered by the low-pass filter.

Further, the input signal frequency equal to IF should also be attenuated to avoid hole in the frequency range. The low-pass filter should also filter this frequency.

In summary, we can say that for a single-band RF spectrum analyzer, we should choose an IF above the highest frequency of the tuning range, make the LO tunable from the IF to the IF plus the upper limit of the tuning range, and include a low-pass filter in front of the mixer that cuts off frequencies below the IF.

Increasing the resolution

Frequency resolution is the ability of a spectrum analyzer to separate two input sinusoids into distinct responses. To separate closely-spaced signals, some spectrum analyzers have IF bandwidths as narrow as 1 kHz, 100 Hz or even 10 Hz. Such narrow filters are difficult to design at a high centre frequency (e.g. 3.6 GHz). So, additional mixing stages are added, typically two to four, to down-convert from the first to the final IF. Figure 8.12 shows an architecture with four mixing stages.

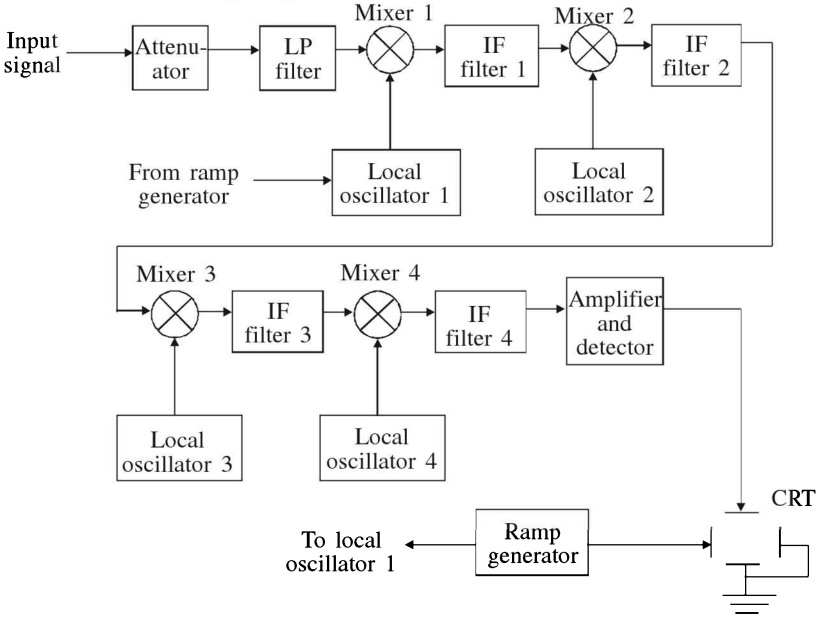


Figure 8.12 Spectrum analyzer with four mixing stages.

The tuning equation for this architecture is

$$f_{sig} = f_{LO1} - (f_{LO2} + f_{LO3} + f_{LO4} + f_{IF4})$$

However $(f_{LO2} + f_{LO3} + f_{LO4} + f_{IF4})$ is chosen to be equal to f_{IF1}.

Considering again the same example of designing the analyzer with tuning frequency of 0 to 2.9 GHz, the following typical frequencies of each block can be chosen:

LP filter cut-off frequency = 3 GHz, f_{LO1} = 3.62 to 6.52 GHz, f_{IF1} = 3.6214 GHz, f_{LO2} = 3.3 GHz, f_{IF2} = 321.4 MHz, f_{LO3} = 300 MHz, f_{IF3} = 21.4 MHz, f_{LO4} = 18.4 MHz, f_{IF4} = 3 MHz.

Note that $(f_{LO2} + f_{LO3} + f_{LO4} + f_{IF4})$ = 3.3 GHz + 300 MHz + 18.4 MHz + 3 MHz = 3.6214 GHz = f_{IF1}.

The narrowest filter in the chain determines the overall bandwidth, and in this example, the filter is in the 3 MHz IF. It is easy to design a filter of 3 MHz with very small bandwidth.

Note that unless two signals are far enough apart, the traces they make fall on top of each other and look like only one response. Fortunately, spectrum analyzers have selectable resolution (IF) filters, so it is usually possible to select one that is narrow enough to resolve closely spaced signals.

Manufacturer's data sheets indicate resolving power by listing the 3 dB bandwidths of the available IF filters. This number tells us how close equal-amplitude sinusoids can be, and still be, resolved. In this case, there is nearly a 3 dB dip between the two peaks traced out by these signals as shown in Figure 8.13.

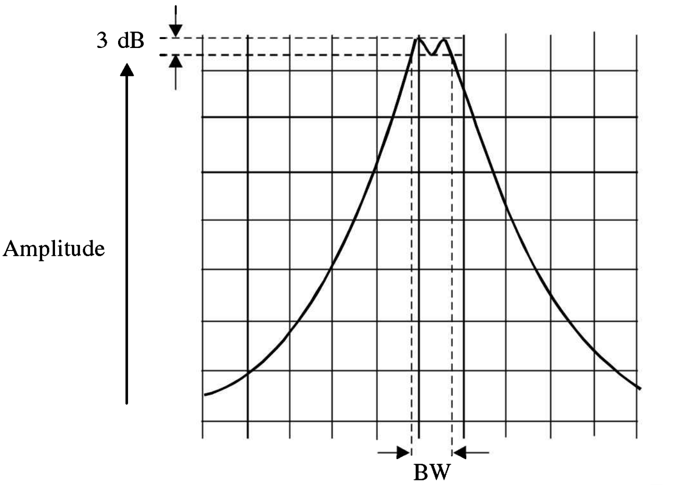


Figure 8.13 Two equal amplitude sinusoids separated by the 3 dB bandwidth (BW) of the selected IF filter.

Some spectrum analyzers use digital techniques to realize the narrower resolution-bandwidth filters of a few Hz. The output from the detector is fed to an analog-to-digital converter (ADC). Digital representation is used with the Fast Fourier Transform algorithm to analyze the signal which is then displayed on the screen.

Frequency instabilities

Two types of frequency instabilities (long term and short term) affect the resolution of the spectrum analyzers. These are basically due to the stability of local oscillator, specially the first one. The first type is the long-term instability which is due to the frequency drift of local oscillator. This will move the spectrum across the screen and if it is too fast, it is difficult to keep the display at the centre.

The second type is the short-term instability (phase instability) of local oscillator that can be observed as phase noise (also called sideband noise). All oscillators are frequency- or phase-modulated by random noise to some extent. Any instability in the local oscillator is transferred to any mixing products resulting from the local oscillator and input signals. So the local oscillator phase-noise modulation sidebands appear around any spectral component on the display that is far enough above the broadband noise floor of the system (see Figure 8.14). The amplitude difference between a displayed spectral component and the phase noise is a function of the stability of the local oscillator. The more stable the local oscillator, the farther down the phase noise. The amplitude difference is also a function of the resolution bandwidth. If we reduce the resolution bandwidth by a factor of ten, the level of the phase noise decreases by 10 dB.

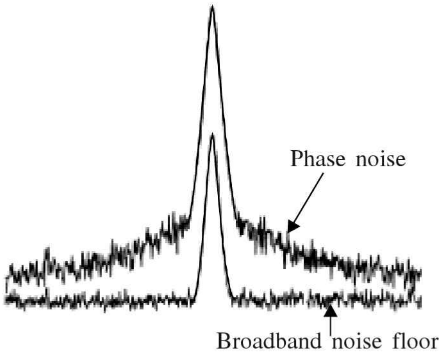


Figure 8.14 Phase noise above the noise floor.

Phase noise is specified in terms of dBc or dB relative to a carrier. It is sometimes specified at a specific frequency offset; at other times, a curve is given to show the phase noise characteristics over a range of offsets. Phase noise becomes the ultimate limitation in an analyzer's ability to resolve signals of unequal amplitude. The phase noise can cover up the smaller signal.

Dynamic range

Dynamic range is the ability of an analyzer to measure harmonically-related signals and the interaction of two or more signals; for example, to measure second- or third-harmonic distortion or third-order intermodulation.

Note that the input mixer of a spectrum analyzer is a nonlinear device and so always generates distortion of its own. It is nonlinear to translate an input signal to the desired IF. But the unwanted distortion products generated in the mixer fall at the same frequencies as do the distortion products we wish to measure on the input signal. So, we can define dynamic range as the ratio, expressed in dB, of the largest to the smallest signals simultaneously present at the input of the spectrum analyzer that allows measurement of the smaller signal to a given degree of uncertainty.

Dynamic range due to internal distortion is a function of the input signal level at the input mixer. Figure 8.15 shows how the level of harmonic and intermodulation components change with the change in the level of fundamental frequency (or frequencies). Change in level of second harmonic component is two times the change in fundamental level and change in third harmonic and third order intermodulation is three times the change in fundamental level.

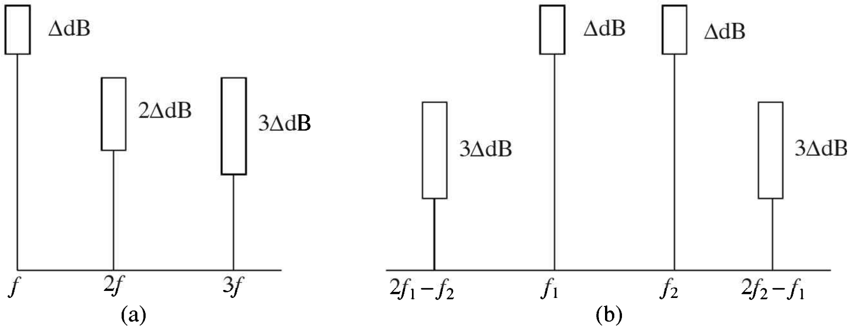


Figure 8.15 Change in the level of (a) second and third harmonics and (b) intermodulation components with change in fundamental frequency (frequencies).

It means that the third-order intermodulation increases at a much more rapid rate than the increase in the two signals. Eventually the intermodulation product will have the same level as the two input signals. This level of equality is called the *third-order intercept point* and is the indication of the upper limit of the dynamic range. This point no doubt is theoretical only, since this much distortion is never acceptable. However, this point can be used to calculate the third-order products. The level of the third-order product, P_3, is given by:

$$P_3 = 3P_{\text{in}} - 2I_{\text{P}} = \text{MDS} \tag{8.3}$$

where

P_3 = level of the third-order product (dBm), P_{in} = power of the two input signals (dBm), I_{P} = power level of third-order intercept, and MDS = minimum detectable signal (dBm).

MDS is essentially the spectrum noise level. From Eq. (8.3), the dynamic range of the spectrum analyzer is the difference in level between the minimum detectable signal and the input that produces a spurious signal equal to the MDS, or

$$P_{\text{in}} - \text{MDS} = \frac{2}{3}(I_{\text{p}} - \text{MDS}) \tag{8.4}$$

For example, the dynamic range of the spectrum analyzer with a third-order intercept point of 20 dBm and a noise level of 80 dBm from the Eq. (8.4) is given by

$$\text{Dynamic Range} = \frac{2}{3}\,[20 - (-80)] \approx 66.67 \text{ dB.}$$

Applications of spectrum analyzer

Some of the common applications are given below:

- Observing purity of a signal
- Analyzing modulated signals
- Studying harmonic components of a signal
- Finding intermodulation content

Limitations of spectrum analyzer

(a) They can measure only steady state signals, so they are impractical for measuring transient events.
(b) We cannot get information about relative phases of the frequency components. The instrument displays power magnitude only—a scalar quantity.
(c) Measurement times are often 10 to 100 times longer than those with the FFT techniques.

8.5 FFT ANALYZER

The Fast Fourier Transform (FFT) is based on the conversion of a time domain waveform to the frequency domain. Joseph Fourier developed the Fourier transform, which converted continuous time domain signals into continuous frequency domain information. The frequency domain information includes magnitude and phase values. However, the FFT analyzer samples discrete points over a certain time interval in the time domain and then records the points digitally. Because the time domain waveform is stored as discrete values, the waveform is not continuous and cannot be converted to the frequency domain using the standard Fourier transform. Instead, another version of the Fourier transform, the Discrete Fourier Transform (DFT), is used to convert the discrete time domain waveform into a discrete frequency domain spectrum.

For a sampled signal, the DFT equation is

$$X_k = \sum_{n=0}^{N-1} x_n \exp \frac{-j2\pi nk}{N}$$

$$0 \le k \le N - 1$$

where

x_n = input discrete time sequence on the interval $0 \le n \le N - 1$

X_k = DFT of x_n

This gives sampled frequency domain representation of the signal with points taken at $f = kf_s/N$, where f_s is the sample frequency to generate original time series.

Because the DFT uses discrete values to transform the time domain waveform into the frequency domain, the frequency domain is divided into discrete frequency components, or *bins*. The number of frequency bins over a given range, called the frequency resolution, is dependent on the number of discrete time domain samples, N, taken. Thus, in order to boost frequency resolution, the number of samples must increase. A limiting factor here is the execution time. The more samples that are taken to increase the resolution, the longer it takes to compute the DFT. The number of mathematical operations to compute DFT is proportional to N^2, where N is the number of samples in the transform.

The FFT is simply a faster method of calculating the DFT. It requires that the time-domain waveform contain a number of samples equal to some power of two. The number of math operations to compute FFT reduce to a number proportional to $N.\log_2 N$. So FFT algorithms are faster than DFT.

If N is not a power of 2, zero padding is done to the end of the sequence to make it a power of 2. The addition of zeros to the end of the time-domain waveform does not affect the spectrum of the signal. However it increases the speed of the algorithm and also frequency resolution is increased since the number of samples are increased ($\Delta f = f_s/N$).

The FFT analyzer consists of a number of components, each of which is necessary for frequency analysis. First, signals must pass through attenuators or gain stages to adjust the amplitude of the signal to match the input range so as to maximize the resolution of the signal. The signal is passed through an anti-aliasing filter to avoid aliasing. Next, a sampler takes discrete points of the analog waveform at a given frequency. This is followed by an analog-to-digital converter (ADC), which converts the samples into digital data. The data is transformed to the frequency domain using the FFT algorithm. The analyzer can use dedicated DSP circuitry to perform the FFT, or the data can be transferred to a PC where the FFT is calculated using software. Figure 8.16 shows the block diagram of a dedicated FFT spectrum analyzer.

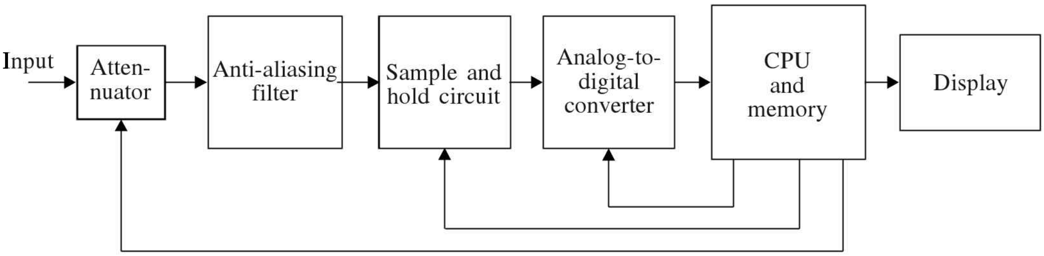


Figure 8.16 FFT spectrum analyzer.

The FFT algorithm is calculated from time records. Each record contains a finite number of discrete points. The FFT converts these time-domain points into a finite number of frequency- domain points. The frequency-domain representation is symmetric about the 0 Hz frequency point. Only the upper-half of the transform, which pertains to 0 Hz and above, is kept. Thus, only a certain number of frequencies can be analyzed. The frequency resolution, or spacing between the bins, must be such that the acquired frequencies are represented properly. In order to increase the frequency resolution the number of samples must increase. This is because the frequency resolution is inversely proportional to the length of the time record.

Whether an FFT analyzer is considered real time or not, is determined by the computing time of the FFT. If the time to compute the FFT is faster than the rate at which the data was sampled, then an FFT can be performed for each point. However, this requires computational power beyond that of most instruments. A shortcut is to take a certain number of samples and transfer them to an intermediate buffer. The FFT takes the data from this buffer to perform its operation. If the FFT is computed before the next time record is placed in the buffer, then the analyzer is running in real time.

The demands of good linearity and high resolution limit the ADCs used in FFT analyzers to those with relatively low sample rates-from a few hundred Ksamples per second to several Msamples per second. This in turn, limits the frequency range of these instruments to about a few kHz, with some capable of a few megahertz. Deep memory allows these instruments to have frequency resolutions of better than 1 mHz. Because of the FFT analyzer's low frequency resolution, we can use FFT analyzers for analysis of vibration and acoustics. We also can use them when testing low-frequency electronics (including control systems) and low-frequency communications.

Dynamic range

The dynamic range represents the difference in level between the greatest signal that can be measured without overload and the smallest signal (quantization noise) that can be displayed together with the larger signal. It can be represented mathematically in dB as follows:

$$\text{Dynamic range} = 20 \log_{10} 2^N$$

For example, the dynamic range for a 10-bit ADC is about 60 dB.

Windowing

As we have seen, the FFT analyzer samples the data for a fixed amount of time. Therefore the resultant spectrum determination represents a spectrum of a periodic function repeated infinitely. The sample represents a window and the data is considered as a periodic function where the data in the window is repeated. The shape of the window affects the spectrum to a degree depending on the type of waveform being analyzed.

Some common window functions are:

- Rectangle
- Hanning
- Hamming
- Kaiser

Table 8.1 defines the sampled values, w_n, of each window function for values of n between $-m$ and $+m$, giving $(2m + 1) = N$ sample points in total. The function I_0 is a *zero-order Bessel function of the first kind.* This function can be computed by means of a power series expansion to any required accuracy. The value of the window parameter α determines the amount of taper of the Kaiser window. With $\alpha = 0$, the window is a flat, rectangular shape; as α increases, the window becomes more strongly tapered.

Table 8.1 Sampled values of window functions

Window type	*Function*
Rectangular	1
Hanning	$0{\cdot}5 + 0{\cdot}5 \cos [n\pi/(m+1)]$
Hamming	$0{\cdot}54 + 0{\cdot}46 \cos (n\pi/m)$
Kaiser	$I_0(\alpha[1 - n^2/m^2]^{1/2})/I_0(\alpha)$

Figure 8.17 shows the shapes of window functions. Because of sudden discontinuities, the rectangular window produces maximum distortion. This window is generally used for transients. For sine waves and periodic functions, other windows are used.

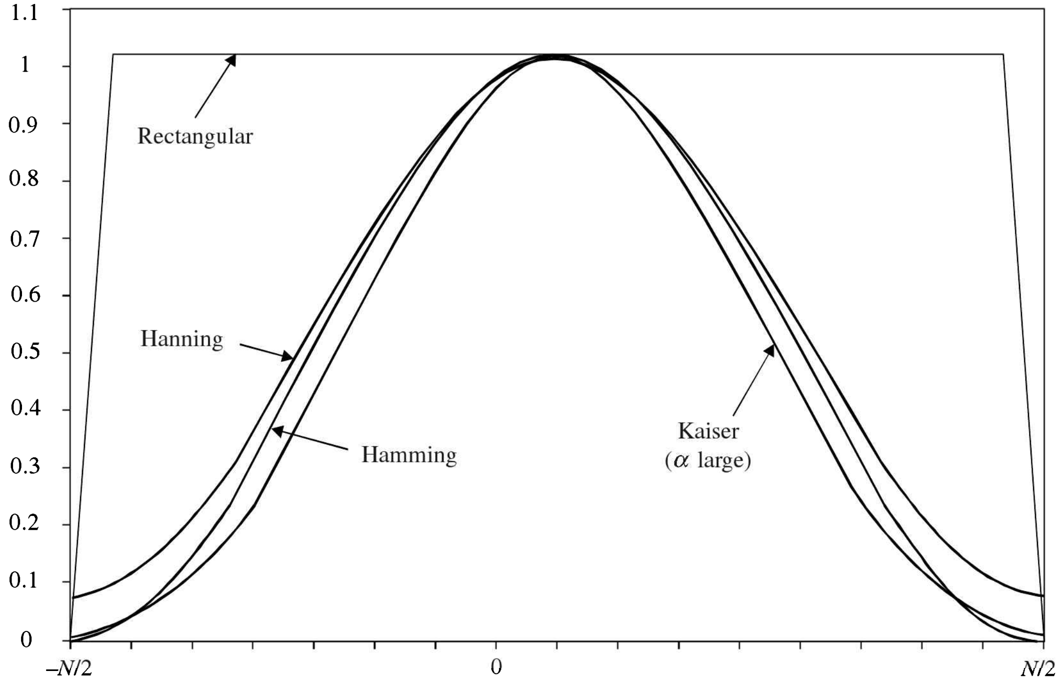


Figure 8.17 Shapes of window functions.

Other features

FFT analysis also lets us see the relative phases of a signal's spectral components. When we measure a transient event such as an impulse, we can use the instrument to calculate the complex transfer function of the circuit. Some analyzers even calculate the transfer function's poles and zeros. Phase information also lets us display the data in more useful formats, such as a Nyquist plot.

Because FFT analyzers digitize and store signals, they also can show how a signal's frequency content changes over time. FFT analyzers can give us that third dimension (time) by generating a waterfall display (Figure 8.18), which shows the results of many acquisitions at once. Each trace gets offset slightly from the previous one, generating a three-dimensional image that often resembles a waterfall.

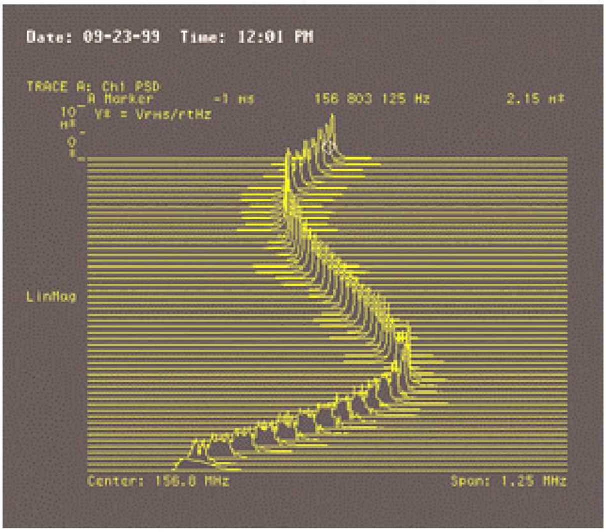


Figure 8.18 Waterfall display.

Further, since computer is used to perform FFT algorithms, other mathematical operations can also be performed on the spectral display. For example, noise effects can be reduced by employing root mean square averaging techniques. This can improve the signal-to-noise ratio performance of the analyzer. However it needs additional time, making the whole process slow.

8.6 VECTOR ANALYZER

FFT analyzers are no good for RF signals. Spectrum analyzers can measure RF signals, but they loose phase information. When we need FFT analysis for RF signals, we can use a vector analyzer. The vector signal analyzer extends the FFT analyzer's capabilities to higher frequencies for narrow band signal analysis. These instruments are ideal for analyzing communications signals. Their capabilities include digital modulation analysis and AM/ FM/ PM demodulation.

A vector analyzer adds a spectrum analyzer front-end to an FFT analyzer (Figure 8.19). The spectrum-analyzer front-end extends the instrument's capabilities into the gigahertz range. As with the spectrum analyzer, a mixer and filter shift the input signal in frequency. In this case, the analyzer shifts the desired input signal's frequency range down to a range compatible with the back-end Fourier analyzer. The filter does not have to be as selective as it is for the spectrum analyzer because its purpose is to prevent aliasing and not to measure the spectral lines. The filter's frequency range can be almost as large as the input frequency range of the Fourier analyzer. As a result, we can use a vector analyzer to view both continuous and transient narrow band RF signals in the frequency domain.

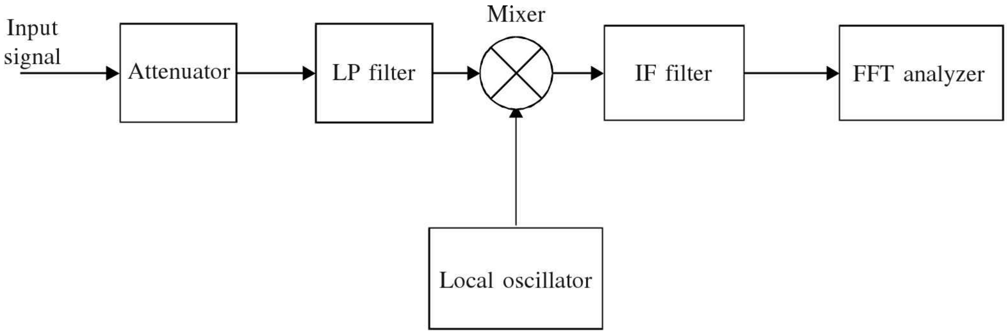


Figure 8.19 Block diagram of vector analyzer.

8.7 DIGITIZING OSCILLOSCOPE

Like vector analyzers, digitizing oscilloscopes (DSOs) (see details in Chapter 3) extend the FFT technique to higher frequencies than those possible with Fourier analyzers. Many DSOs provide basic spectral analysis as a math function. But DSOs do not have the 12-bit or 16-bit amplitude resolution of FFT analyzers. High-speed DSOs typically resolve signals to 8 bits. Still, their high sampling rate (several billion samples per second) lets us see the spectrum of transient events.

Note that when we use a DSO's frequency domain capabilities, small time-domain distortions introduced by the DSO's front-end that are difficult to see in the time domain become obvious in the frequency domain, especially when we view a signal's amplitude in decibels. This is because DSOs are optimized for time-domain analysis. These distortions introduce low-level harmonics into the scope's FFT when we view simple signals such as sine waves.

Many high-sample-rate DSOs achieve their high sample rates by interleaving several fast ADCs per channel. The ADCs vary slightly in linearity, gain, offset, frequency response, and timing. These mismatches, interleaved in time, introduce extraneous signals at multiples of FS/N into the scope, where N is the number of interleaved ADCs and FS is the interleaved sample rate. Except for offset, all of these errors result in mixing of our signal and the FS/N components. For example, assume that four 1-Gsample/s ADCs are interleaved to provide 4 Gsamples/s. The FS/N components are at 1 GHz and 2 GHz. The offset differences between the ADCs result in spectral components at 1 GHz and 2 GHz. With a 100 MHz sine wave input, timing errors result in fictitious spectral components at 900, 1100, and 1900 MHz.

High-speed scopes have 8 bits of resolution as opposed to the 12 bits or more in FFT analyzers. Fewer bits result in more quantization errors. With fewer bits than FFT analyzers and vector analyzers, DSO's quantization noise can add nonlinearities to a signal. The DSO's noise floor is higher by 20 to 40 dB, and the nonlinearities of the ADC process are correspondingly higher than in an FFT analyzer. Nonlinearities contaminate the FFT by allowing the quantization noise to mix with the signal, generating spurious spectral components instead of random noise. Many DSOs add a controlled *noise source* to the signal before quantization and subtract it after quantization. This *dithering* is designed to optimize the frequency content of the signal and minimize the effects of nonlinearities in the ADC, including the quantization process.

Whenever possible we can use the DSO's real time capture mode as opposed to the equivalent-time sampling mode, for spectrum analysis. The FFT assumes that the samples are equally spaced, so just a few picoseconds of sampling jitter can destroy the FFT's signal-to-noise ratio (SNR). The samples from an equivalent-time record represent many independent acquisitions. Timing inconsistencies in the circuit from repetition to repetition and DSO trigger jitter easily result in tens of picoseconds of effective jitter, making the FFT useless. If we must use an equivalent-time mode, set the number of averages, N, as high as possible. This will statistically reduce the jitter by the square root of N.

The high sample rate of DSOs lets us view the spectrum of high-speed transients as well as continuous signals.

Table 8.2 summarizes the benefits and limitations of each equipment used for spectral analysis.

Table 8.2 Benefits and limitations of equipment for spectral analysis

Instrument	*Benefits*	*Limitations*	*Applications*
Spectrum Analyzer	• Widest frequency range • Widest frequency span	• Requires a stable signal • Scalar measurements • Does not go down to dc • Large minimum frequency resolution	• RF spectral plots • EMI scans
FFT Analyzer	• Narrow frequency resolution • Vector measurements • Measurements down to dc • Transient analysis • Rich analysis capabilities	• Narrow frequency range • Narrow frequency span	• Vibration analysis • Acoustic analysis
Vector Analyzer	• Benefits of FFT analyzer • Widest frequency range	• Does not go down to dc	• RF spectral plots with phase information
Digital Oscilloscope	• Narrow frequency resolution • Widest bandwidth for transient analysis • Measurements down to dc	• Narrow frequency span • Lower dynamic range • Aliasing • Spurious signals • Limited analysis features	• Spectral plots of transient signals • Phase analysis of transient and continuous signals

8.8 LOGIC ANALYZER

This analyzer is used for different applications as compared to spectrum and FFT analyzers. Suppose we need to look simultaneously at the inputs and outputs of a 16-bit counter to determine a timing error and we only have a 2-channel scope. How do we look at them all? Let us say we have just developed timing diagrams for a board, full of digital circuitry. How do we verify them? Suppose an intermittent glitch appears to occur on one of the data lines of the microcontroller system, causing the processor to get incorrect data. We cannot trigger on just the glitch with the analog oscilloscope. What do we use to capture and analyze it?

With the wrong tool, solving these kinds of problems can be very time consuming. The best tool is the logic timing analyzer. The following section presents the basics of the logic analyzer.

8.8.1 Oscilloscope or Logic Analyzer

When given the choice between using an oscilloscope and using a logic analyzer, we normally prefer an oscilloscope because a scope is more familiar to us. It is one of the most *general purpose* of all electronic instruments. However, it has some shortcomings that limit its usefulness in some applications. A logic analyzer may yield more useful information in many of these applications. However, since a logic analyzer is *tuned* to the digital world, it does not have as broad a use as an oscilloscope. Because of some overlapping capabilities between a scope and a logic analyzer, either may be used in some cases. How do we determine which of the two instruments better for our application? By analyzing the following questions:

When should we use a scope?

We should use an oscilloscope when we need to see small voltage excursions on our signals and when we need high time-interval accuracy.

Generally, an oscilloscope is the instrument to use when we need high vertical or voltage resolution and to measure high-time interval resolution. That is, it can measure the time interval between two events (Figure 8.20).

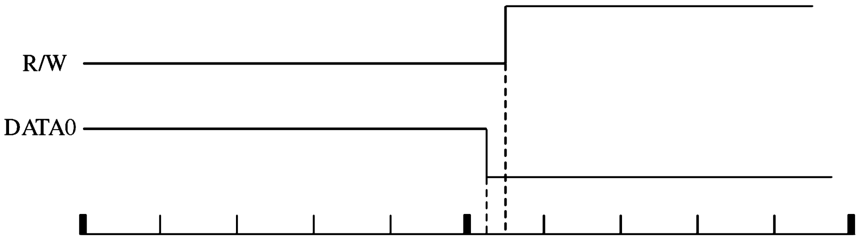


Figure 8.20 Time interval between two waveforms.

To describe it another way, if we need to see every little voltage excursion, we need a scope. Many scopes, including the new generation digitizing ones, can also provide very high time-interval resolution. That is, they can measure the time interval between two events with very high precision. Overall, an oscilloscope should be used when we need parametric information about our signal.

When should we use a logic analyzer?

We should use a logic analyzer when we need to see (i) the relationships between many signals at one glance, (ii) the signals in our system the same way our hardware does, and (iii) when we need to trigger on a pattern of HIGHs and LOWs on several lines and see the result.

Logic analyzers grew out of oscilloscopes. They present data in the same general way that a scope does; the horizontal axis is time, and the vertical axis is logic HIGH or LOW and not voltage. A logic analyzer does not provide as much time-interval accuracy as the oscilloscope. It can capture and display eight or more signals simultaneously, something that scopes cannot do. A logic analyzer reacts the same way as our logic circuit does when a single threshold is crossed by a signal in our system. It will recognize the signal to be either LOW or HIGH. It can also trigger on patterns of highs and lows on these signals. So when do we use a logic analyzer? When we need to look at more lines than the oscilloscope can show us, provided we can manage without precise time-interval information. If we need to look at every little transition on the waveform, a logic analyzer is not a good choice.

Logic analyzers are particularly useful when looking at time relationships or data on a bus, e.g. a microprocessor address, data, or control bus. It can decode the information on microprocessor buses and present it in a meaningful form. Generally, when we have passed the parametric stage of design, and are interested in timing relationships among many signals and need to trigger on patterns of logic highs and lows, the logic analyzer is a good choice.

8.8.2 Types of Logic Analyzers

Figure 8.21 shows how a logic analyzer is interfaced with a system under test (SUT).

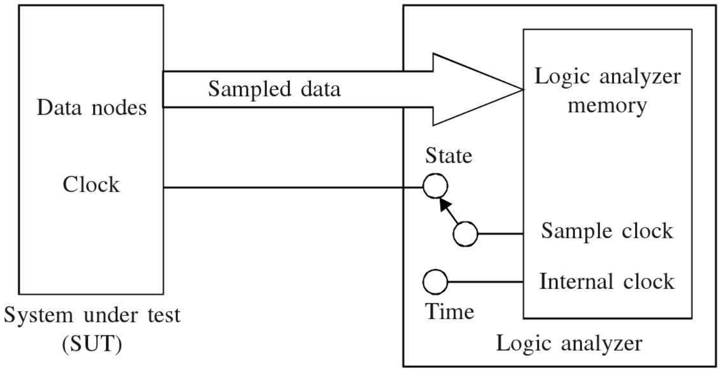


Figure 8.21 Connection of a logic analyzer to a system under test (SUT).

On a clock pulse, information on the many nodes probing the SUT is latched into the logic analyzer memory as a 1 or a 0 depending upon whether it was greater or less than the threshold voltage. The type of analyzer depends on the clock source as follows:

- If the clock is taken from the SUT, it is called state/synchronous/software analyzer.
- If the clock is generated within the analyzer, it is called *timing analyzer*. Since its clock is not synchronized to the SUT, it is also known as asynchronous analyzer.

The logic analyzer captures its data as a series of 0s and 1s. Current analyzers include post processing that translates these binary patterns into more meaningful displays. In the case of a state analyzer the information is disassembled back into microprocessor mnemonics and even the original high level source code, while for a timing analyzer the display or trace is a timing diagram.

8.8.3 Logic Timing Analyzer (LTA)

A timing analyzer is the part of a logic analyzer that is analogous to an oscilloscope. The timing analyzer displays information in the same general form as a scope, with the horizontal axis representing time and the vertical axis as logic levels of HIGH and LOW. Because the waveforms on both instruments are time-dependent, the displays are said to be in the *time domain*.

Logic timing analyzer's sampling method

A logic timing analyzer works by sampling the input waveforms at the clock frequency to determine whether they are HIGH or LOW. It cares about only one user-defined voltage-threshold. If the signal is above threshold when it samples, it will be displayed as a 1 or HIGH by the analyzer. By the same criterion, any signal sampled that is below threshold is displayed as a 0 or LOW. From these sample points, a list of ones and zeros is generated that represents a one-bit picture of the input waveform. As far as the analyzer is concerned, the waveform is either HIGH or LOW, and there are no intermediate steps. This list is stored in memory and is also used to reconstruct a one-bit picture of the input waveform, as shown in Figure 8.22.

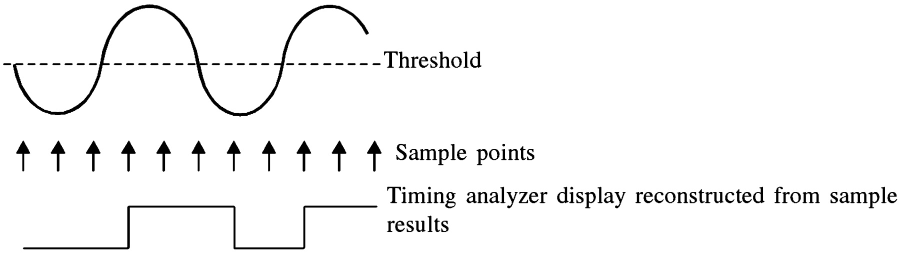


Figure 8.22 Sampling of the input waveform.

This tendency to square everything up would seem to limit the usefulness of a timing analyzer. We should remember, however, that it is not intended as a parametric instrument. If we want to check the rise time of a signal with an analyzer, this is the wrong instrument. But if we need to verify timing relationships among several lines by seeing them all together, a timing analyzer is the logical choice.

For example, imagine that we have dynamic RAM in a system that must be refreshed every 2 ms. To ensure that everything in memory is refreshed within that 2 ms, a counter is used to count up sequentially through all rows of the RAMs and to refresh each row. If we want to make certain that the counter does indeed count up through all rows before starting all over, a timing analyzer can be set to trigger when the counter starts and to display all of the counts. Parameters are not of great concern here; we merely want to check that the counter counts from 1 to *N* and then starts all over again.

When the timing analyzer samples an input line, it is either HIGH or LOW. If the line is at one state (HIGH or LOW) on one sample and the opposite state on the next sample, the analyzer *knows* that the input signal transitioned at sometime in between the two samples. It does not know when, so it places the transition point at the next sample, as shown in Figure 8.23. This presents some ambiguity as to when the transition actually occurred and when it was displayed by the analyzer. Worst case for this ambiguity, also called horizontal resolution, is one sample (clock) period, assuming that the transition occurred immediately after the previous sample point.

With this technique, however, there is a trade-off between resolution and total acquisition time. Remember that every sampling point uses one memory location. Thus, the higher the resolution (faster sampling rate), the shorter the acquisition window.

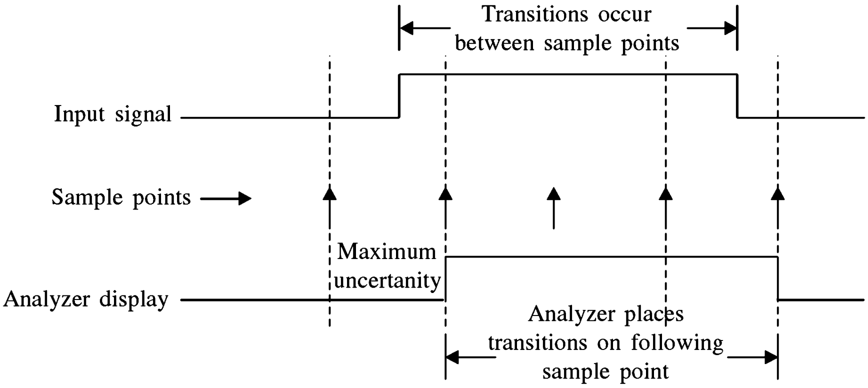

Transitions occur between sample points
Input signal
Sample points
Maximum uncertanity
Analyzer display
Analyzer places transitions on following sample point

Figure 8.23 Transition in the analyzer.

Transitional sampling

When data is captured on an input line with data bursts, as shown in Figure 8.24, sampling rate should be adjusted to high resolution (e.g. 4 ns) to capture the fast pulses at the beginning. This means, however, that a timing analyzer with 4K memory would stop acquiring data after 16.4 μs. And the second data burst cannot be captured.

Note that data was sampled and stored for a long time even when there was no activity an the input side. This uses memory without providing additional information. In transitional sampling, sampling is done only when transitions occur. This uses memory more efficiently.

To accomplish this a *transition detector* can be used at the input of the timing analyzer together with a counter. The timing analyzer will now store only those samples which were

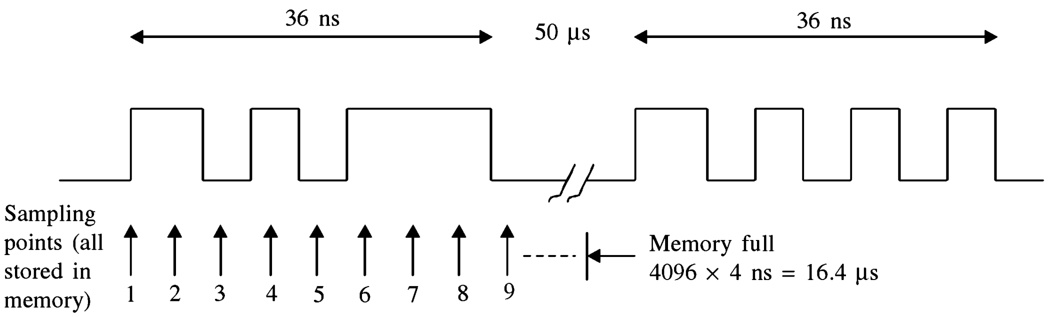


Figure 8.24 Transitional sampling.

preceded by a transition together with the elapsed time from the previous transition. Thus only memory locations per transition are used and no memory is used if there is no activity at the input. This technique is called *transition timing*. With this, in the example of Figure 8.24, not only the second burst, but also the third, fourth and fifth burst can be captured depending on how many pulses per burst are present. At the same time, timing resolution of 4 ns can be maintained. This is shown in Figure 8.25. Only 18 memory locations are required (14 sampling points plus 14 time intervals.

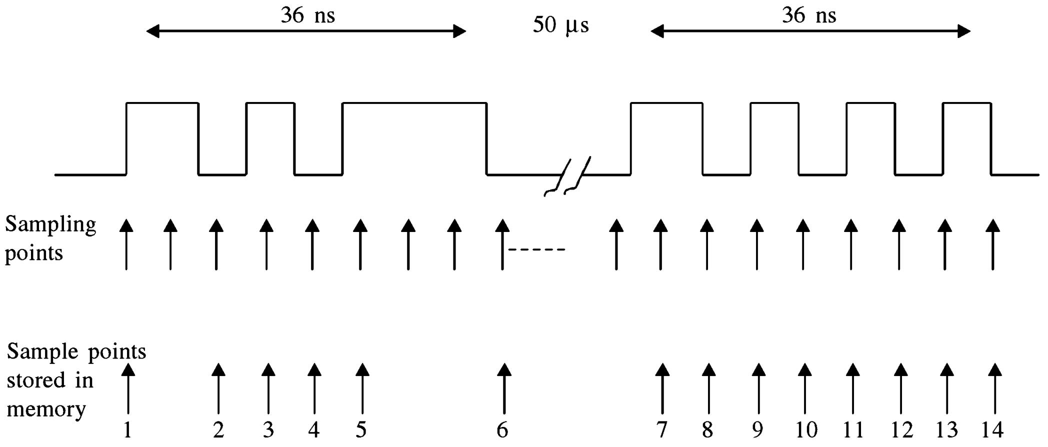


Figure 8.25 Transitional sampling using transition detector.

Effective memory depth

This equals the total time data that is captured divided by the sampling period. For example, if the total time data captured is 33 μs at a sampling period of 4 ns, the effective memory depth will be 75KB.

Capturing narrow pulses

The logic analyzer resolution is equal to the clock period. If the input signal is a pulse that has a duration less than the resolution, the pulse will either be captured and displayed as a pulse with a duration equal to the sample clock period or it will be missed entirely (see Figure 8.26).

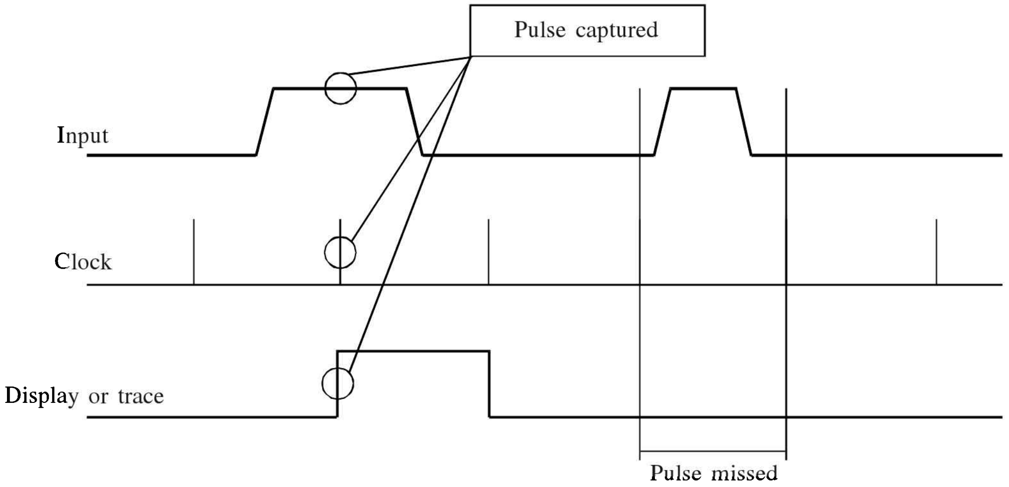


Figure 8.26 Missing of a pulse.

In the first case of the analyzer, clock occurs while the pulse is active, so this pulse is captured and displayed as a minimum width sample. As shown, the trace goes HIGH on the clock edge where the input was sampled HIGH and remains HIGH until the next clock edge. Due to the analyzer resolution, the true pulse width could have been anywhere between 0 to 2 clock periods. In the second case, the pulse occurs between clock samples, so it is missed with a standard logic timing analyzer.

There are two solutions to this dilemma. One solution is to include pulse stretching circuitry in series with the input (Figure 8.27). Such circuitry will be set by any input transients and not reset until after the next sample point. All spikes are captured as pulses of width identical to the sample period.

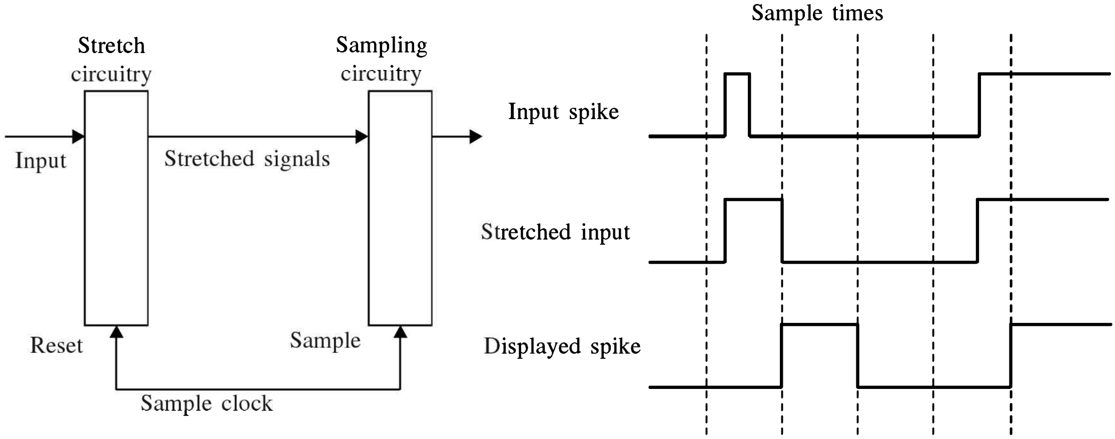


Figure 8.27 Including pulse stretching circuitry.

The second solution is to include the glitch detection circuitry.

Glitch capture. One problem with digital systems is the infamous *glitch*. Glitches have a nasty habit of showing up at the most inopportune times with the most disastrous results. How do we capture a glitch that occurs once every 36 hours and sends our system into the weeds? Once again the timing analyzer comes to the rescue!

A glitch can be caused by capacitive coupling between traces, power supply ripples, high instantaneous current demands by several devices, or any number of other events. High performance DSO's provide a glitch trigger mode where the user can specify a trigger on a pulse of less than or greater than a specified pulse width. However, since a timing analyzer samples the incoming data and can keep track of any transitions that occur between samples, it can readily recognize a glitch (Figure 8.28). In the case of an analyzer, a glitch is defined as any transition that crosses logic threshold more than once between samples.

The analyzer already keeps track of all single transitions that occur between samples, as already discussed. To recognize a glitch, we *teach* the analyzer to watch for multiple transitions and trigger on them.

It is helpful to have the ability to trigger on a glitch and display data that occurred before it. This can help us to determine what caused it. This capability also enables the analyzer to capture data only when we want it, i.e. when the glitch occurred. Think about the example we mentioned in the beginning paragraph of this section. We have a system that crashes periodically because a glitch appears on one of the lines. We could set up a trigger condition that captures the crash and then look at the control lines before the crash occurred. Another alternative is to use an analyzer without glitch trigger capability and sit in front of the machine, waiting, until we see the glitch. Unfortunately, neither of the above are practical alternatives. If we can tell the analyzer to trigger on a glitch, it can stop when it finds one, capturing all the data that happened before it. We let the analyzer wait and when the system crashes, we have a record of what led up to the error.

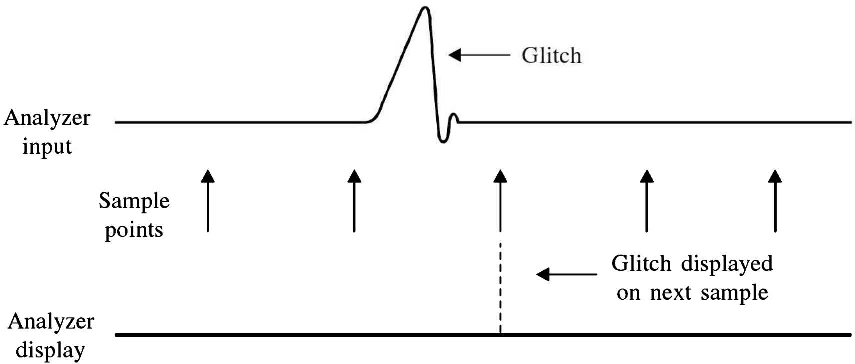


Figure 8.28 Capturing a glitch.

Another use of the glitch display is to detect situations where the waveform might be undersampled. A display that contains many glitches might be undersampled. This can be tested by rerunning the experiment at a faster sweep speed. If the glitches are greatly reduced, it was undersampled.

Note that the definition of a glitch depends upon the sample period. If the sample period is one minute, then a 30 second pulse may be detected as a glitch.

Logic timing analyzer skew

Skew is the difference in delay between the channels on a logic timing analyzer. Its effect will be important when trying to resolve multiple channels of high frequency data. Logic analyzer manufacturers have devoted significant efforts to keep the skew of their probes to a minimum. Consider Figure 8.29. Note that the horizontal axis has been expanded for clarity, in practice a display would consist of many hundred samples.

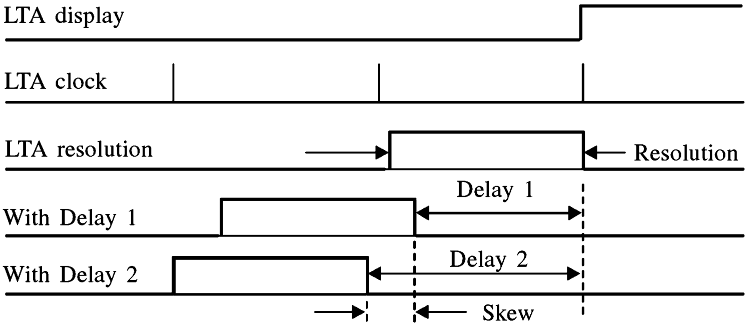


Figure 8.29 Timing analyzer skew.

The first waveform shows the start of a pulse as displayed by a logic timing analyzer. Due to the analyzer resolution, all that the user can say is that the pulse started somewhere between the (last) two clock pulses. That is, the uncertainity of the reading equals the analyzer resolution. This is the situation at the analyzer sampling circuit. Prior to the sampling circuit there are probes going to the system under test (SUT). These probes introduce delays. With one channel the delay is not important. With multiple channels what is important is the difference in the delays or the skew. The uncertainty of the reading equals the resolution plus the skew of the probes.

Consider an analyzer with a 200 MHz sampling clock and probes with a skew of 2 ns. The uncertainity will be (5+2) ns. If the trace is identical for two channels it is possible for the first channel to have changed within 7 ns of the second. Generally this resolution is not required. However if it is, then the analyzer should be used to trigger an oscilloscope. An indication of the skew may be found by observing the same signal with multiple probes as illustrated in Figure 8.30.

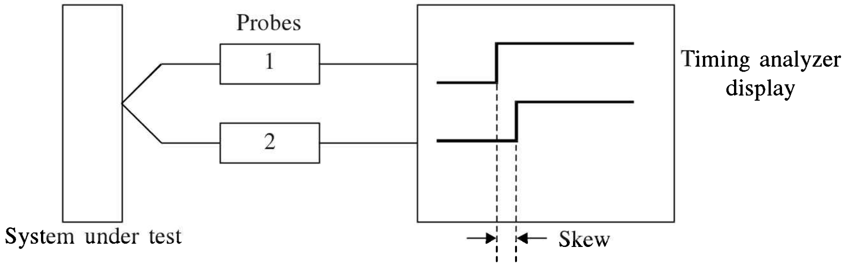


Figure 8.30 Observing the skew.

Triggering the timing analyzer

Another term that should be familiar to oscilloscope users is *triggering*. It is also used in logic analyzers, but is often called *trace point*. Unlike an oscilloscope which starts the trace right after the trigger, a logic analyzer continuously captures data and stops the acquisition after the trace point is found to display the data. Thus a logic analyzer can show information prior to the trace point, which is known as *negative time*, as well as information after the trace point.

Pattern (state) triggering

Setting trace specifications on a timing analyzer is a bit different from setting trigger level and slope on an oscilloscope. Many analyzers trigger on a pattern (state) of highs and lows across input lines. (see Figure 8.31). In this example the trigger is set to the state 1,1. The input data is continuously rippled through the analyzer memory. Once the trigger condition is satisfied then, in the trigger starts trace mode, the memory is filled once more and its contents frozen. These contents are then displayed as the resultant trace. Other modes such as trigger ends trace or trigger centres trace are also available.

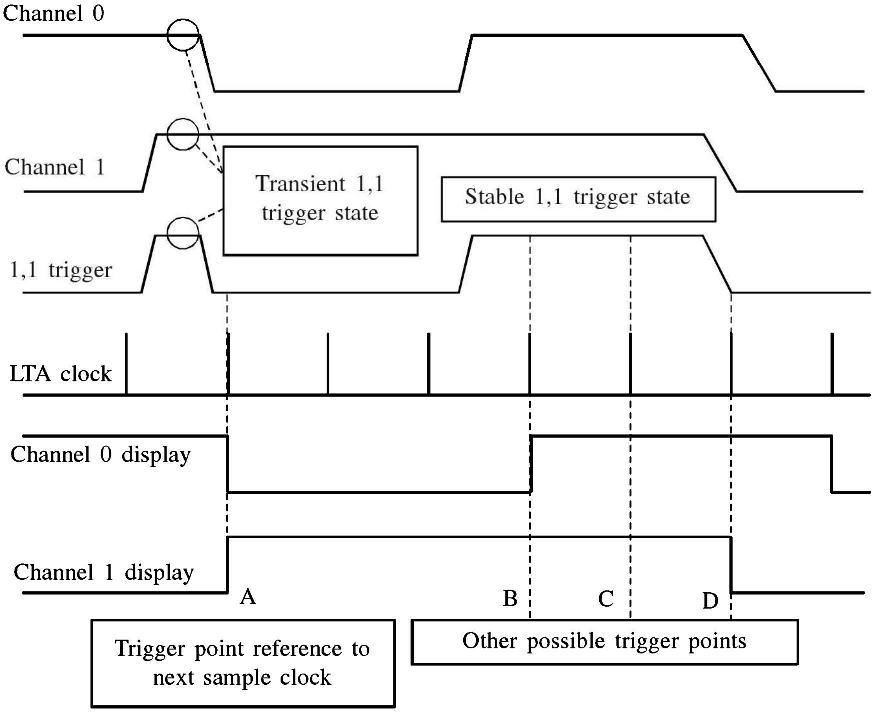

Channel 0
Channel 1
Transient 1,1 trigger state
Stable 1,1 trigger state
1,1 trigger
LTA clock
Channel 0 display
Channel 1 display
A
B
C
D
Trigger point reference to next sample clock
Other possible trigger points

Figure 8.31 Triggering of the timing analyzer.

To the triggering example two refinements are normally added. One is the trigger duration filter where the trigger conditions must be satisfied for up to or less than a specified time to be considered a valid trigger condition. The greater than may be used to eliminate the possibility of triggering on transient states. If the trigger duration filter is set to a short interval, triggering at clock edges A and B is possible, and lengthening the duration filter would change the trigger point to clock edge C or D.

An important feature of a logic timing analyzer is that the trigger circuitry is in parallel with the data capture circuitry. In this example, the trigger conditions were satisfied between the first and second sampling points. At the second sampling point the analyzer will know that the triggering conditions have been satisfied so the trace will commence. This is the situation illustrated where the state 1,1 did occur between the sample points, so clock edge A is a trigger point. Note that the trace shows information leading up to the trigger point, that is the logic analyzer displays data in negative time.

The second triggering refinement is the ability to trigger on a glitch. This is illustrated in Figure 8.32, where the analyzer triggers on the state 1,1 plus a glitch on channel 0.

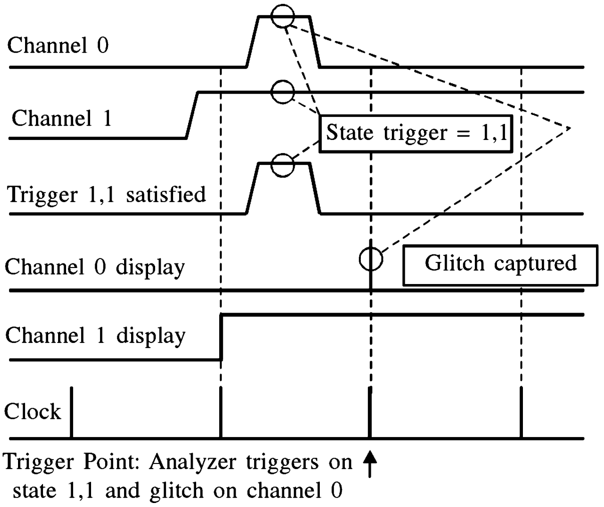


Figure 8.32 Triggering on a glitch.

In using an analyzer one critical consideration is to find a unique condition that can be used to reference the activity of interest. It is possible that no unique trigger can be found at the activity of interest. One possibility is to use the trigger plus the delay facility. As shown in Figure 8.33, the analyzer is triggered on the unique single pulse but the trace is delayed sufficiently to capture the area of interest.

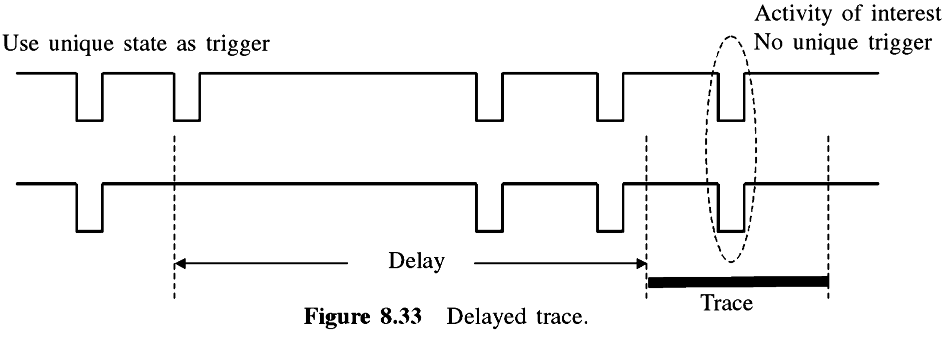


Figure 8.33 Delayed trace.

Edge triggering

Edge triggering is a familiar concept to those accustomed to using an oscilloscope. When adjusting the "trigger level" knob on a scope, we could think of it as setting the level of a voltage comparator that tells the scope to trigger when the input voltage crosses that level. A timing analyzer works essentially in the same way on edge triggering except that the trigger level is preset to logic threshold. Why include edge triggering in a timing analyzer? While many logic devices are level-dependent, clock and control signals of these devices are often edge-sensitive. Edge triggering allows us to start capturing data as the device is clocked. As a simple example, take the case of an edge-triggered shift register (Figure 8.34) that is not shifting data correctly. Is the problem with the data or the clock edge? In order to check the device, we need to verify the data when it is clocked on the clock edge. The analyzer can be told to capture data when the clock edge occurs (rising or falling) and catch all of the outputs of the shift register. Of course, in this case, we would have to delay the trace point to take care of the propagation delay through the shift register.

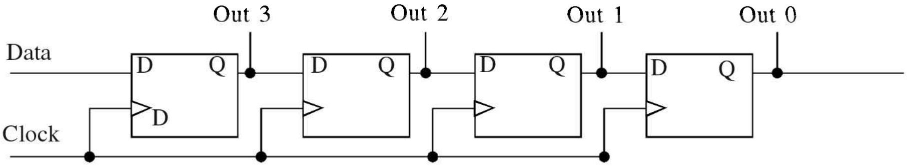


Figure 8.34 A shift register.

8.8.4 Logic State Analyzer (LSA)

The logic state analyzer is used to capture the state sequence flow of the system under test (SUT). The most important requirement is that the logic analyzer should capture the information exactly the same as the SUT. This implies that the clock to the logic analyzer is the same as the reference clock of the SUT. For example, if the SUT samples/interprets the information on the falling edge of the clock then this is the signal that must be used by the logic analyzer. Using the wrong signal will cause subtle errors that may not be detected. Due to the problems of probing, most analyzer manufacturers produce special preprocessor modules that also handle the clock requirements. With a preprocessor the selection of the clock is not an issue for the user.

Logic state analyzer displays

The most simple state analyzer display is a binary list of 1s and 0s. This form of display was used in the early analyzers. The first refinement was to display the trace in other formats such as octal, decimal, or hexadecimal.

While hexadecimal displays may have been tolerated in the smaller projects, they are not very user friendly and inverse assemblers that translated the 1s and 0s into processor specific mnemonics became standard with logic state analyzers. The example given in Figure 8.35 is a portion of a trace with a microprocessor. It is the hexadecimal address and data information that is captured by the analyzer. The mnemonics etc. are generated by the

```
   Listing including the RTS

   Line     Addr     Data     Mnemonic/Operand
   001      C185     20       JSR      C41E
   002      C187     1E                operand
   003      01FD     C1                internal
   004      01FD     C1                write      <---- return address
   005      01FC     87                write      <---- to stack
   006      C187     C4                operand
   007      C41E              Subroutine
   ...
   ...
   047      C48F     60         RTS
   048      C490     8D                prefetch
   049      01FB     70                internal
   050      01FC     87                read       <---- return address
   051      01FD     C1                read       <---- from stack
   052      C187     C4                internal
   053      C188              Next Instruction
|<--- Hardware -->|<----- LSA software ---->|<--- User Interpretation
--->|
```

Figure 8.35 Display of logic state analyzer.

software within the analyzer using the data and status information collected by the analyzer hardware.

This example is (vertically) divided into three parts:

1. The first group lists the hexadecimal information as captured by the logic analyzer hardware, this shows exactly what is executing on the microprocessor busses. Note that the logic state analyzer captures **all** the microprocessor bus activity. Therefore, the trace gives both the program fetch and the program execution.

2. The second group is the logic analyzer's software interpretation of what is happening. Thus in line 001 the logic analyzer has captured 20H which the microprocessor would interpret as a Jump to Subroutine. The logic analyzer has been programmed to interpret the data exactly the same way as the microprocessor under test, so it displays the mnemonic JSR. The actual operand or destination is read from later samples.

3. Layered on the above are the pieces of information added by the user, for example, the interpretation of the stack write and read.

If the logic analyzer has the original source code available it is possible to translate the target addresses into labels or the assembly level mnemonics back into the high level language source code. Thus in the example, lines 1 through 6 would be displayed as a function call, while lines 47 through would correspond to the end of the function.

One of the most important requirements of a logic state analyzer is the ability to capture the information of interest.

Trace or display qualifiers

Trace or display qualifiers are used to:

- Filter out states that are not necessary for the measurement
- Limit the size of the analyzer memory necessary.

Display and trace qualifiers (generally) display the same results on the screen, the difference is that the trace qualifier picks out the information of interest before storing it in the analyzer memory, whereas the display qualifier or filter reads only the information of interest from memory.

Examples of qualifiers might be where the analyzer is requested to:

- Trace/display only memory writes. This might be used to determine how a block of data is being changed.
- Trace/display only op code fetches. This would be used to trace the program flow. In the example above, the user would read that the program goes from address C185 to C41E through to C48F and then return to C188 (see Figure 8.35).
- Trace/display only program reads (the execution of the instructions is not captured). In the previous example, only lines 1, 2, 6, and 7 would be captured.
- Trace/display only the address range. This might be used to trace only a specified function.
- Trace/display not address range. This might be used to eliminate capturing, and filling the logic analyzer memory with activity from monitor or library routines.

Example. "Trace only memory writes". Let the logic analyzer be formatted to trace only writes to the address Serial_Port which in this case is the physical address 8003H. In this example, the trace shows that the byte 41H was written to the address 8003H followed by 42H, 43H, etc.

```
Addr    Data    Mnemonic/status
8003     41       memory  write    3.256  s
8003     42       memory  write       17  µs
8003     43       memory  write    1.040  ms
8003     44       memory  write    1.048  ms
8003     45       memory  write    1.042  ms
8003     46       memory  write    1.044  ms
```

Time tagging

All logic state analyzers include a time tag. This is very useful, especially with a qualified trace. In the example above, once equilibrium is reached, the trace indicates that the data is written every 1.04 ms. (For a serial device this corresponds to 9600 baud).

Triggering the logic state analyzer

One feature that determines the quality, and hence the cost of a logic state analyzer is its triggering capability; the ability of users to specify trigger conditions that enable them to window the trace about the activity of interest. Let us consider an example program which transmits a message to a serial I/O device. The basic program consists of two loops (Figure 8.36). The outer loop is executed once for each character transmitted while the inner loop is executed many hundreds of times waiting for the I/O device to become ready for the next character.

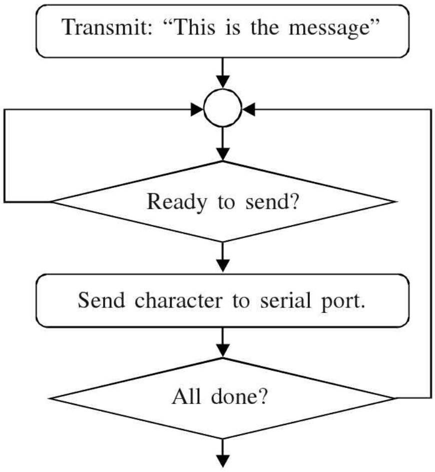

Transmit: "This is the message"
Ready to send?
Send character to serial port.
All done?

Figure 8.36 Flow chart.

Program Listing:

```
Label     Mnemonic   Operand   Comments and program explanation.
START     LDX        #MESSAGE  ;Initialize index pointer
          LDB        ,X+       ;number of bytes to transmit at
                                start of message
NEXT      LDA        ,X+       ;character to send
          JSR        TRANSMIT  ;send the character
          DECB                 ;one less character to send
          BNE        NEXT      ;loop until all done
CONTINUE
          .........
          .........
TRANSMIT  PSHS       A         ;save accumulator A
TRAN1     LDA        STATUS    ;poll status waiting
          BPL        TRAN1     ;clear to send is bit 7
          PULS       A         ;return character to send
          STA        IOPort    ;send it
          RTS
MESSAGE   DB         19,"This is the message."  ;the message
IOPort    EQU        8000H     ;define locations
STATUS    EQU        8001H
```

The listing shows the code implemented using an 8-bit microprocessor. The code fragment contains two program loops. In the main code, the B accumulator is initialized with the number of bytes of the message. Each time a character is sent, B is decremented until it reaches 0 when the outer loop (and code fragment) should exit. That is, it should exit when the string is sent. For the second or inner loop the status of the I/O device is continually polled until it is ready to transmit the character read by the main loop.

Without a trigger condition the logic state analyzer will capture what is happening at the time the "run" command is given. Usually the user will be interested in a specific piece of code. That is, after the "run" command is given the analyzer should wait until a specified condition is detected and then reference the trace to that (trigger) condition.

Sequential triggering

In many problems the designer will be interested in what happens at a certain point in the program after some previous conditions have occurred. In the sample program the question might be what occurs on the output port after certain data is read from the data string. The desired trace condition would be of the form

"find data value then trace after/about address output_port".

This is known as sequential triggering and is a feature contained in all better analyzers. For the simple example program given, it will offer no advantage over the other triggering modes, but with programs that contain several paths it will allow the capability of determining what happened when the software reached a certain point after transiting only the specified program paths.

For the example program shown in Figure 8.37, it is desired to trace the activity at point C but only when the program has come via path A and path B. Further, since it is possible

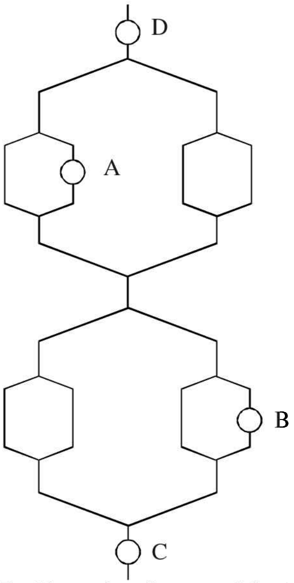


Figure 8.37 Example of sequential triggering.

that the trigger condition A is satisfied on one pass and trigger condition B satisfied latter, which is not what is required, the sequence trigger logic should then be reset when the state D occurs. The complete triggering specifications will be:

"find in turn state A then state B then trace on state C reset on state D.

The trigger condition shown illustrates two sequential levels. State-of-the-art logic state analyzers contain many levels of sequential triggering.

Don't care trigger

The don't care trigger is most useful when it is not known in what area of the program the microprocessor is executing. Often if there are problems, the first measurement will be to take a trace of what is happening at that point of time. Observing the result, the designer can determine what code is actually being executed. Once the code being executed is known, it is usually possible to make some assumptions as to the program flow to that point. If necessary, the logic state analyzer may be used to verify those assumptions by setting the trigger conditions to points in the path. If the trace shows activity in an unexpected area of memory, then the state analyzer, in the trigger-ends-trace mode, may be used to determine how the microprocessor reached the problem code.

8.8.5 Block Diagram of Logic Analyzer

Figure 8.38 shows the block diagram of a logic analyzer.

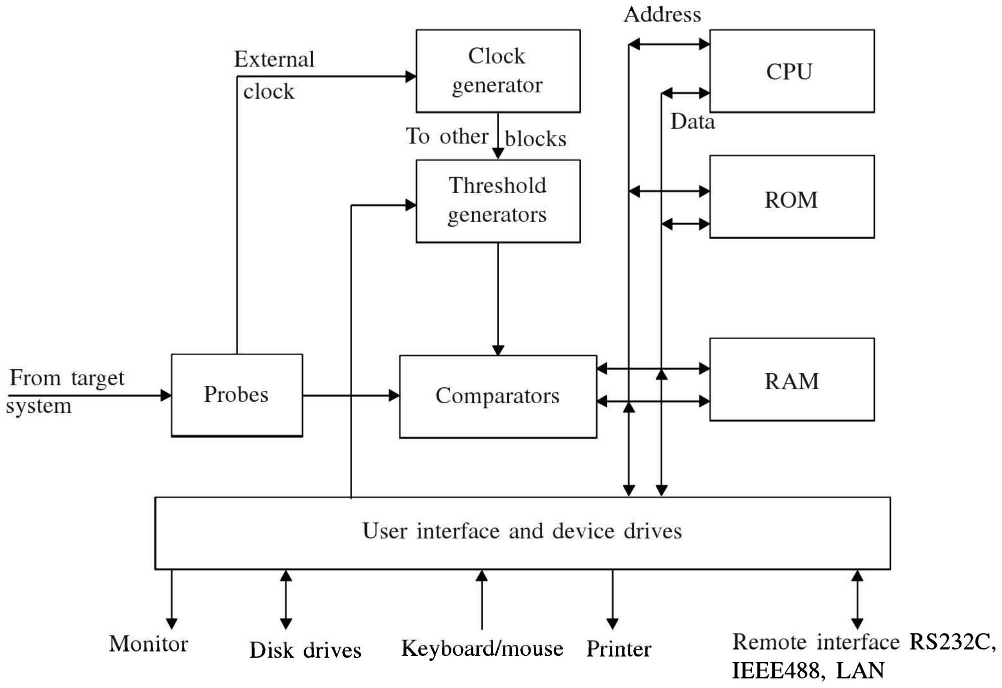


Figure 8.38 Block diagram of a logic analyzer.

8.8.6 Interfacing Target System

Probes are used to interface target systems which are to be analyzed. Because of a large number of input channels, it is not practical to use unique cables to connect individual channel. Instead 8 or more channels are grouped into pods (plug-on-devices). Each pod connects to the system as a single cable. Passive probes are most common. Typically, each channel is terminated with 100 kΩ and 8 pF. These probes being of small size can be terminated right at the point of connection to the target system. Thus, additional stray capacitance due to the wires from large active pods to the circuit under test are avoided.

Many a times, a probe has a probe adapter or a preprocessor. A preprocessor provides not only fast, reliable and correct mechanical connection to the target system but also necessary electrical adaptation like clocking and demultiplexing to capture the system's operation correctly. Some microprocessors prefetch information from memory that may never get being executed. They can also distinguish prefetched information from executed. Preprocessors also come with inverse assembler to decode HEX information to microprocessor mnemonics.

The high-density probe is a next-generation design that incorporates a sophisticated mass termination technology to deliver four times as many connections in the same area as traditional square-pin style probes (Figure 8.39).

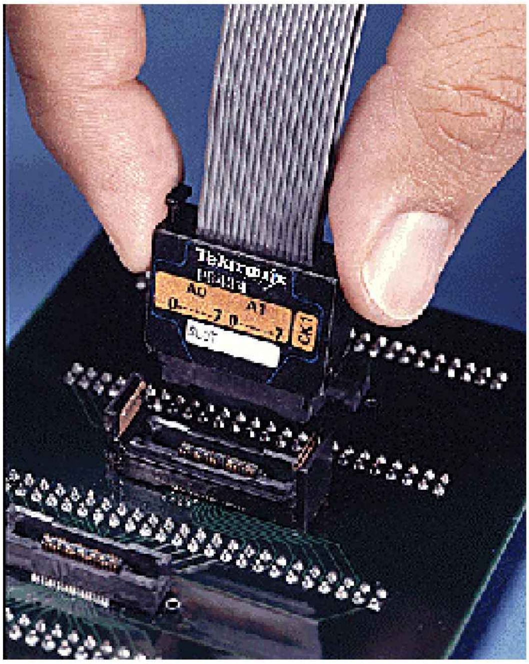


Figure 8.39 High density probe.

This high-density solution offers outstanding electrical characteristics for superior dc and ac loading of high-speed signals. Its capacitance is a mere 2 pF with a controlled impedance environment to minimize reflections. This solution also provides complete isolation between channels and shielding to prevent outside interference. All these features add up to one thing —faithful transmission of high-speed timing and state information from the design to the logic analyzer.

To ensure a reliable connection, this probe also includes an optional latching capability that only adds a fraction of an inch to the size of the connector. This positive latch keeps the connector from accidentally coming loose with up to ten pounds-plus of pull. And an easy release mechanism makes it simple to disconnect.

Although standard probing meets the demands of existing fixtures, this high-density approach will be the probing solution of choice for the future, especially for emerging processor-based designs. Modern processors have hundreds of pins located over extremely small areas. Attaching a logic analyzer to these chips today can literally take hours with the end result weighing several pounds and consuming considerable board real estate.

Besides providing outstanding signal quality, the high-density connector, with its 34 signal connections, considerably streamlines the complexity of attaching hundreds of channels. It takes up just a quarter of the space of the square-pin approach for the same amount of connections—a mere 1.09″ × 0.39″ for 34 channels. For example, to completely connect to a complex processor like the Pentium® requires only four of these high-density assemblies.

Condensing the probe size results in unprecedented probing flexibility. Stub lengths can be drastically reduced and hundreds of through holes in the test fixture eliminated, keeping the impact on signal quality to a minimum. The probe also dramatically simplifies attaching the logic analyzer to the design, with the potential of misconnections almost eliminated. It is easier to access and cool high-speed processors while the probes are attached.

8.8.7 What to Look for in a Logic Analyzer?

The speed limit

It is absolutely essential that a logic analyzer be able to keep pace with the fastest circuitry in our design. To ensure clear visibility into the inner workings of the design, a logic analyzer must be able to capture and analyze its fastest events. That means for today's high-speed processors, the logic analyzer should deliver timing resolution in the sub-nanosecond region. In addition, precise setup and hold windows are needed to support high-speed state acquisition. Although most logic analyzers today are equipped to handle 100 MHz state speeds, they are not able to keep up with next wave of high-speed general-purpose processors.

Speed alone, however, does not make for success. A logic analyzer must also intelligently sift through millions of pieces of data to quickly find the information we need. That is why triggering—how a logic analyzer determines when to capture data—is so important. When searching for a problem in a sophisticated design, the visible symptoms are often an intricate web of interrelated hardware and software events. As a result, a logic analyzer must be able to monitor a multitude of signals and search for any number of possible combinations of software instructions, data access, timing violations, glitches, or setup and hold violations. Traditional triggering, that only makes one if-then-else decision per clock, often fails to locate the subtle, asynchronous events that typically threaten the integrity of advanced designs.

The channel count

Applications today are not only getting faster, they are growing in size. High-end processor designs already require 100 to 150 channels to cover all the input and outputs of the device. In complex applications, such as a multiprocessor design, the channel count soars into the hundreds. There are even designs that exist today that require monitoring over 1000-plus signals to fully analyze the application. So even if our design does not require hundreds of channels now, we may want a logic analyzer that can easily expand to accommodate future design demands.

It is equally important that each and every one of those channels be capable of sustaining the logic analyzer's top acquisition speed. Unfortunately, most logic analyzers only provide a limited number of high-speed channels even when they support hundreds of connections. But designs no longer just have a handful of high-speed signals. It is not uncommon to need to view every input and output with sub-nanosecond timing resolution.

A logic analyzer's probe adapters and disassembly software is another crucial factor in how effectively you can analyze a complex processor with hundreds of connections. With so many processor families on the market, a logic analyzer must support literally scores of probe adapters. And the disassembly software should give a clear view of which instructions are actually executed and which conditional branches are actually taken.

Deep memory

Untangling the relationship between tens or even hundreds of signals over many cycles requires that the logic analyzer extract and store an extraordinary amount of data. Although most logic analyzers offer deeper memory capabilities, it usually comes at the expense of some other function. Only by shutting down an important feature, such as time stamping, can some logic analyzers deliver the memory depth required for advanced debugging. Such compromises in acquisition capabilities heavily dilute the benefits of longer memory.

Complexity

If we keep focused on the three basic requirements of speed, width and depth, we will quickly zero-in on the right logic analyzer for our design requirements. Although it is important to determine up front which of those criteria is the most important, remember there is a good chance that our next design will need a logic analyzer that gives us exceptional performance in all three areas.

QUESTIONS

8.1 Explain harmonic distortion, total harmonic distortion (THD), THD + n, SINAD and IMD.

8.2 Draw the block diagram of distortion analyzer and explain its operation. What steps are taken to reduce hum and noise in the analyzer?

8.3 Draw the block diagram of wave analyzer and explain its operation. Why is it called non real time analyzer?

8.4 Draw the block diagram of SMPTE intermodulation analyzer and explain its operation and various features.

8.5 Draw the block diagram of CCIT intermodulation analyzer and explain its operation and various features.

8.6 What is meant by spectrum analysis? Compare it with time domain analysis.

8.7 What are the applications of spectrum analyzer? Why is it called real-time analyzer?

8.8 Draw the block diagram of filter bank spectrum analyzer and explain its operation and various features.

8.9 Explain the construction and operation of each block of a Superheterodyne Spectrum Analyzer. What will be the typical frequencies selected for various blocks for an input range of 2.9 GHz?

8.10 Draw the block diagram of a superheterodyne spectrum analyzer. Explain its operation. How is its resolution improved? What are its limitations?

8.11 (a) Explain resolving power, frequency instabilities and dynamic range of a spectrum analyzer.

(b) Calculate the dynamic range of a spectrum analyzer with a third order intercept point of 20 dBm and a noise level of 100 dBm.

8.12 Explain the construction and operation of an FFT spectrum analyzer.

8.13 Explain the concept of Windowing in the FFT spectrum analyzer. Draw the shapes of a few window functions.

8.14 Compare the FFT analyzer with the superheterodyne spectrum analyzer.

8.15 What is a vector analyzer? Draw its block diagram and explain its operation. Compare it with the FFT analyzer.

8.16 (a) Explain the operation of logic timing analyzer and logic state analyzer.

(b) Give specific applications where one is preferred over the other.

8.17 Draw the block diagram of a logic analyzer and explain its operation.

8.18 How are target systems interfaced with the logic analyzer?

8.19 What are important parameters to be considered while selecting a logic analyzer?

Chapter 9

Electronic Counters

9.1 INTRODUCTION

Electronic counters are used for the measurement of frequency, period, time interval, phase, event counting and many other related signal parameters. They find widespread applications in laboratories, production lines and service centres of telecommunications, electronic components, aerospace, military, computer, education and other industries. The advent of microprocessors has brought about a proliferation of products to the counter market.

Before counters were invented, frequency was measured with a frequency meter using a tuned device which had very low accuracy. Frequency counter was one of the first instruments employing digital electronics to measure signal parameters more precisely. Electronic counters have normally more than one input channel and provide several measurement functions such as:

- Frequency
- Period
- Time interval
- Totalizer
- Pulse width
- Rise time
- Fall time
- Slew rate
- Duty cycle
- Phase difference
- Frequency ratio
- Velocity

Such counters are called *universal counters*.

In this chapter, we will study the fundamentals of conventional counters, reciprocal counters and time interval counters. The types of measurements they perform and the important considerations that have significant impact on measurement accuracy and performance, will also be explained.

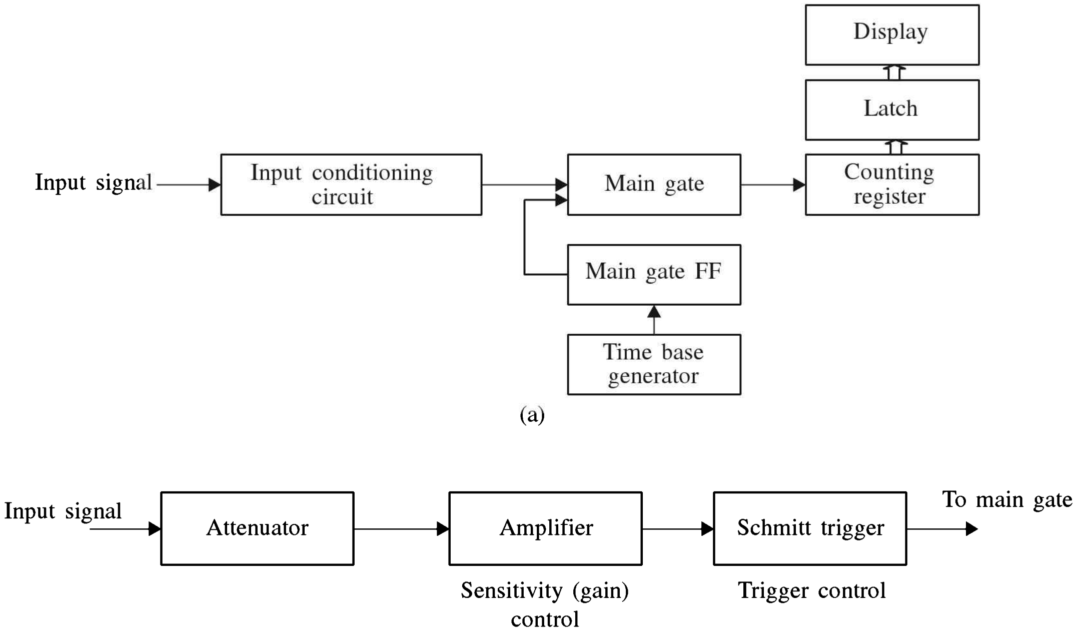
9.2 CONVENTIONAL ELECTRONIC COUNTERS

The conventional electronic counter is a digital instrument which basically measures the frequency of an input signal. It may also be designed to perform related basic measurements like the period of the input signal, ratio of the frequency of two input signals, time interval between two events and totalizing a specific group of events. We will study the operation of the counter for each measurement through its block diagram.

9.2.1 Basic Measurements

Frequency measurement

The frequency, f, of repetitive signals is defined as the number of cycles of that signal per unit of time and is measured by the conventional counter by counting the number of cycles, N, and dividing it by the time interval, t. Figure 9.1 shows the block diagram of the counter in its frequency measurement mode.



(a)

(b)

Figure 9.1 (Cont.)

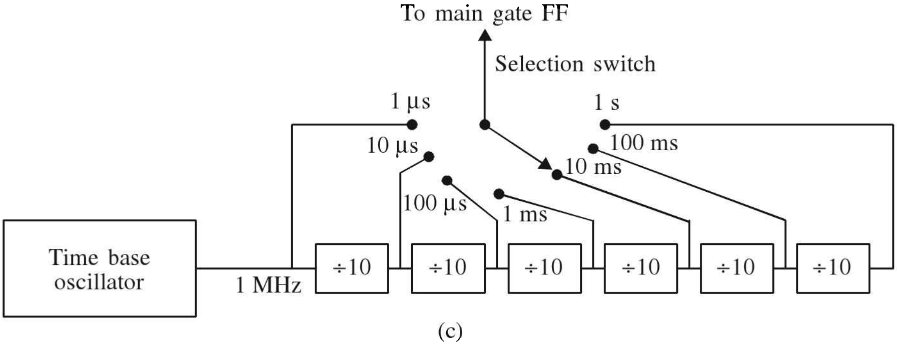


Figure 9.1 Block diagram: (a) frequency counter, (b) input conditioning circuit, and (c) typical time base circuit.

The input signal is first conditioned to form a signal compatible with the internal circuitry of the counting register. The block diagram of this conditioning circuit is given in Figure 9.1(b). It consists of an attenuator, an amplifier and a wave shaping circuit in the form of Schmitt trigger. Attenuator control, gain control and trigger controls are available on the instrument panel which control the attenuation level of the input, gain of the amplifier and threshold levels and hysteresis window of the Schmitt trigger.

The output of the conditioned circuit produces a pulse train which is fed as one input of the main gate. The other input of the main gate is fed from main gate flip-flop (FF) which decides the opening time of the main gate. With the main gate open, pulses are allowed to pass through it and get totalized by the counting register and displayed through the latch. The main gate flip-flop (FF) is controlled by the time base.

The time base consists of a number of decade (÷10) blocks which divide the frequency by 10. The input to the first decade block is provided by the time base oscillator. By dividing this oscillator frequency a number of times by 10, a pulse train with different time is generated. A selector switch is then used to select proper time for the pulse output from the time base. Figure 9.1(c) shows a typical time base circuit generating pulses with period starting from 1 μs to 1 s.

The number of pulses totaled by the counter for the selected gate time yields the frequency of the input signal. It is given by the following expression:

$$\text{Frequency, } f_A = \frac{\text{counts } (N)}{\text{gate time } (t)} = \text{Counts } (N) \times \text{Frequency of the time base } (f_C)$$

For example, if the number of pulses totaled by the counting register is 5000, the selected gate time is 1 s, and the frequency of the input signal is 5000 Hz. If the number of pulses totaled by the counting register is 500, and the selected gate time is 1 ms, the frequency of the input signal is 500 kHz.

It is apparent that the accuracy of the frequency measurement is dependent on the accuracy of the time base. Consequently, most counters employ highly stabilized crystal oscillators.

Additional functions

Normalizing, presetting and prescaling are other functions which are sometimes employed in conventional counters.

Normalizing counters. The normalizing counter displays the frequency of the input signal being measured multiplied by a numerical constant. If f_A is the frequency of the input signal, the displayed value, f_O, is given by

$$f_O = K \cdot f_A$$

where K is a numerical constant (normalizing factor) and is set via thumbwheel switches or by a built-in memory circuit.

This technique is commonly used in industrial applications for measurement of rpm or flow rate.

Preset counters. Preset counters provide an output when the display exceeds the number that is preset in the counter via means such as thumbwheel switches. The output is normally used for controlling other equipment in industrial applications like batch counting and limit sensing for engine rpm measurements.

Prescaled counters. Reliability of high frequency measurement is limited by the input amplifier trigger, the speed of the main gate switches and the counter circuits. One technique that is employed which increases the range of the frequency response without exacting high speed capabilities of the main gate and counter circuits is simply to add a prescaler (divider). The prescaler divides the input signal frequency by a factor, N, before applying the signal to the main gate. This technique is called *prescaling*. (see Figure 9.2). However, the main gate has to remain open N times longer in order to accumulate the same number of counts in the counting register. Therefore, prescaling involves a tradeoff. The frequency response is increased by a factor of N, but so is the measurement time to achieve the same resolution. A slower and less expensive main gate and counter circuits can be used, but at the expense of an additional divider.

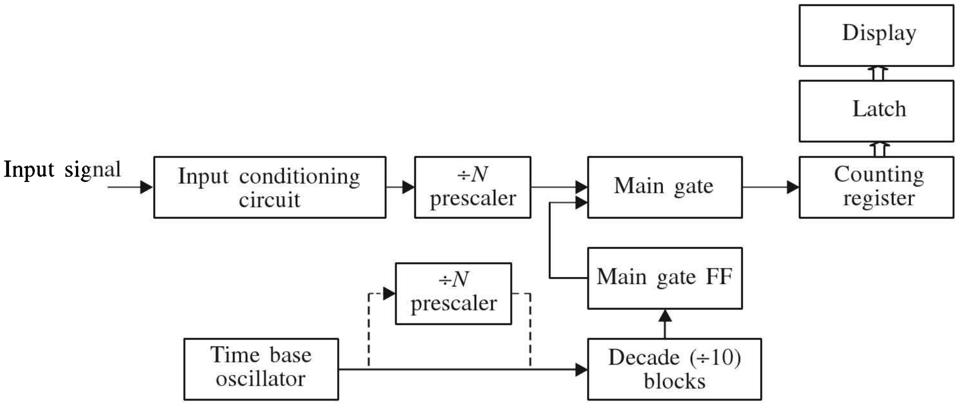


Figure 9.2 Block diagram of prescaled counter.

Period measurement

The period of an input signal is the inverse of its frequency, i.e. it is the time taken for the signal to complete one cycle.

$$\text{Period} = \frac{1}{f_A}$$

If the time is measured over several input cycles, then the average period of the repetitive signal is determined. This is often referred to as multiple period averaging.

The basic block diagram for the conventional counter in its period measurement mode is shown in Figure 9.3. In this mode of measurement, the duration over which the main gate is open is controlled by the frequency of the input signal rather than that of the time base. The counting register counts the output pulses from the time-base for one cycle of the input signal.

$$\text{Period} = \frac{1}{f_A} = \frac{\text{counts } (N)}{\text{frequency of the time base } (f_C)} = \text{Counts } (N) \times \text{Gate time } (t)$$

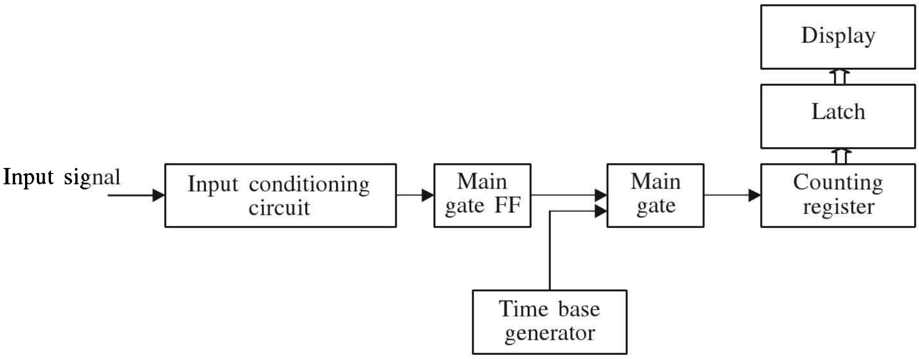


Figure 9.3 Block diagram of frequency counter in period measurement mode.

For example, if the time base is set for 1 μs and if the count is 4625, then the period is equal to 4.625 ms.

Gate can be started (triggered) at +ve or -ve edge of the input pulse and stopped at –ve or +ve edge or started at +ve edge and stopped at the next +ve edge.

The conditioned input signal may also be divided so that the gate is open in decade steps of the input signal period rather than for a single period. This is the basis of the multiple period averaging technique.

Further, lower the frequency of the input signal, the longer the time for which gate is opened and more accurate is the result. We can define the resolution as the smallest time interval that can be measured on the counter and is reciprocal of the time base frequency. Period measurement, therefore, allows more accurate measurement of unknown low-frequency signals because of increased resolution. For example, a frequency measurement of 100 Hz on a counter with 10-digit display and a 1s gate time will be displayed as 0000000.100 kHz. A single period measurement of 100 Hz on the same counter with 10 MHz time base would display 000010000.0 s. The resolution is improved 1000-fold.

Time interval measurement

It is the measurement of time interval between two events or between two points spread in time. This is similar to period measurement. The basic block diagram for the conventional counter in its time interval measurement mode is shown in Figure 9.4.

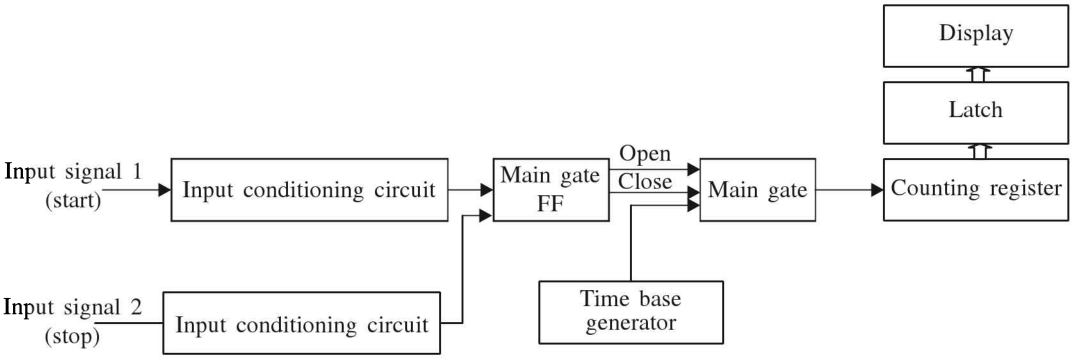


Figure 9.4 Block diagram of frequency counter in time interval measurement mode.

The main gate is controlled by two independent inputs, the START input, which opens the gate, and the STOP input which closes it. Time base (clock) pulses from the time base generator are accumulated for the time duration for which the gate is open. The accumulated count gives the time interval between the START event and the STOP event.

We can also control the gate using triggering level control. The time interval measured is then between different voltage levels of the signal such as those shown in Figure 9.5. The input conditioning circuit must be able to generate the START pulse at the 1.0 V amplitude

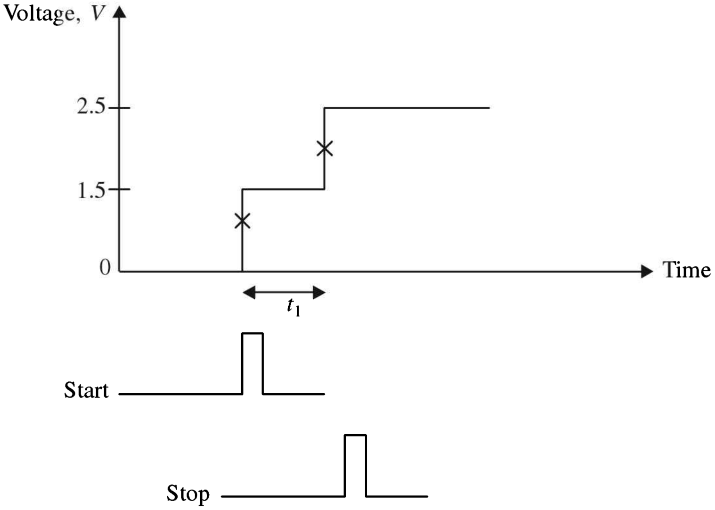


Figure 9.5 Measurement of time interval, t_1, between two levels.

point, and the STOP pulse at the 2.0 V amplitude point. Inherent in the time interval measurement, therefore, is the dual dimensionality, amplitude and time. It is this dimensionality that places much more stringent requirements on the input amplifier/trigger dual than those necessary for the measurement of frequency or period.

Frequency ratio measurement

The ratio of two frequencies is determined by using the lower-frequency signal for gate control while the higher-frequency signal is counted by the counter, as shown in Figure 9.6.

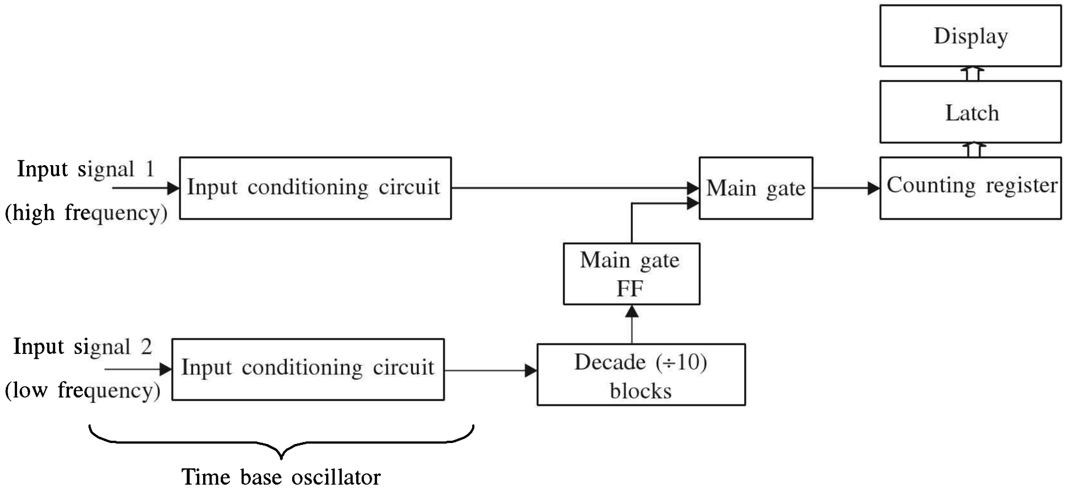


Figure 9.6 Block diagram of frequency counter in ratio measurement mode.

Low frequency signal is divided further by ten a number of times and then the ratio is measured. This is called *multiple ratio measurement*. It gives more accuracy because it gives large count for more time.

Totalizing mode of measurement

Totalizers are used to measure number of events for a specific time. The basic block diagram is same as that of frequency measurement (Figure 9.1). Main gate FF is used to provide START and STOP signals for starting and stopping the totalizing activity. We can also use separate input channels for providing START and STOP signals. Even a manual switch can be used for this purpose.

9.2.2 Design Considerations

Major considerations that affect the performance of various building blocks of the counter are discussed next.

Input considerations

The block diagram of the input conditioning circuit is shown in Figure 9.1(b). The following parameters of the frequency counter depend upon the input conditioning circuit:

- Sensitivity
- Input coupling
- Dynamic range
- Input impedance
- Trigger level
- Trigger slope
- Automatic gain control

Sensitivity. The sensitivity is defined as the minimum specified input signal that can be counted. Sensitivity is usually specified in terms of the rms value of a sinusoidal input. Therefore, for pulse type inputs, the minimum pulse amplitude sensitivity is $2\sqrt{2}$ of the specified value of the trigger level.

The amplifier gain and the voltage difference between the Schmitt trigger hysteresis levels determine the counter's sensitivity. Better sensitivity may not make a counter more better. With a highly sensitive input, noise can cause false triggering. Optimum sensitivity is largely dependent on input impedance. The higher the impedance, the more susceptible is the counter to noise and false counts.

If input to the counter is equal to the input to the Schmitt trigger, we may consider the separation between the hysteresis levels as the peak-to-peak sensitivity of the counter. To effect one count in the counter's counting register, the input must cross both the upper and lower hysteresis levels. This is summarized in Figure 9.7. UTL represents Upper Threshold Level and LTL the Lower Threshold Level of the Schmitt trigger.

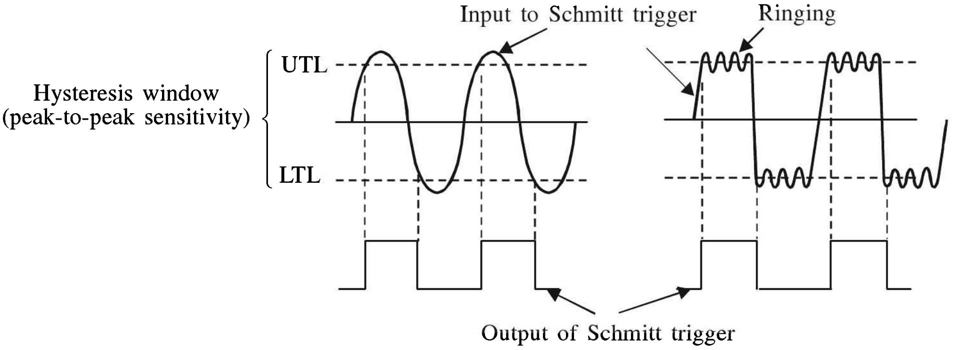


Figure 9.7 Effect of input sensitivity and ringing.

Input coupling. If the input contains a large dc content, then the total signal may shift above the UTL and never produce a counting pulse (see Figure 9.8). Therefore, ac coupling must be provided to avoid the dc level shift in the signal. This works best with the frequency and period measurements and is less useful for time interval measurements because latter are often made on logic signals.

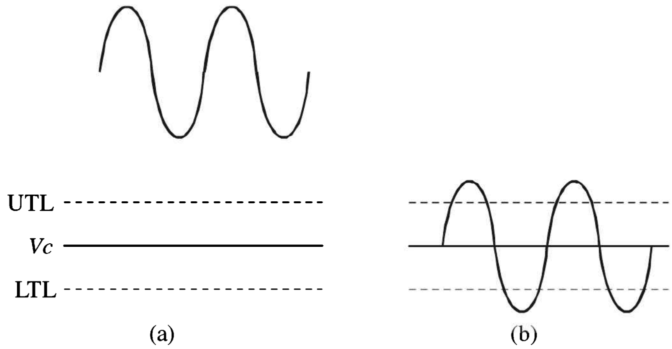


Figure 9.8 Effect of (a) dc and (b) ac input coupling.

Dynamic range. The dynamic range of the input is defined as the linear range of the operation of the input amplifier. However, it is not important for the input amplifier of a frequency counter to be always absolutely linear. Exceeding the dynamic range may not cause false counts but the speed of response may decrease due to reduction in the input impedance and saturation. Protection mechanisms are provided to protect the circuit from large inputs.

The more sophisticated inputs with wide dynamic range usually employ step attenuators with attenuation positions such as X1, X10, and X100.

Input impedance. The input of an amplifier is represented by an impedance consisting of a parallel combination of resistance and capacitance. For example, 1 MΩ resistance in shunt with a 35 pF capacitance. For frequencies up to around 10 MHz, 1 MΩ input impedance is usually preferred which in the majority of sources connected to the input are not loaded. Beyond about 10 MHz, however, the inherent shunt capacity of high impedance inputs rapidly reduces the input impedance. For this reason, 50 Ω impedance levels, which can be provided with low shunt capacity, are preferred. Because of noise and related problems, 20–25 mV values are considered optimum with 50 Ω inputs. A sensitivity of 1 mV is possible at the cost of noise problems.

Trigger level. The function of the trigger level control is to shift the hysteresis levels (UTL and LTL) (keeping the hysteresis window width same) above or below ground to enable positive or negative pulse trains respectively, to be counted.

Many counters provide a three position level control (as shown in Figure 9.9)—a "preset or normal" position, a position normally labeled "+" for positive pulses, and a position normally labeled "–" for negative pulses. The more sophisticated counters provide a continuously adjustable trigger level control, adjustable over the whole dynamic range of the input. This more flexible arrangement ensures that any signal within the dynamic range of the input and of an amplitude consistent with the counter's sensitivity can be counted.

Counters are also equipped with a trigger amplitude control, which allows the window width to be changed.

Trigger slope. Trigger slope control is designed to generate the output pulse from the Schmitt trigger either by a signal with a positive (+) slope (going from one voltage level to another of a more positive level regardless of polarity) at UTL or by a signal with a negative (–) slope at LTL. This is most useful for time interval measurements. It allows the selection of the edge of the input signal to use.

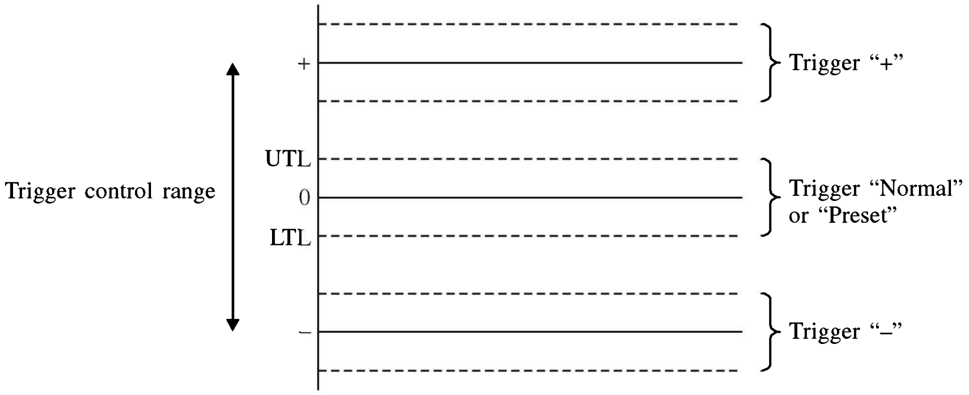


Figure 9.9 Trigger level control.

Automatic gain control. Automatic gain control (AGC) adjusts the gain of the amplifier-attenuator section by the magnitude of the input signal. So it provides automatically adjustable sensitivity control. A trade-off exists between the speed of response of the automatic gain control and the minimum frequency signal that can be counted. For this reason the lower frequency limit for AGC inputs is usually around 50 Hz. AGC inputs, therefore, are useful primarily for frequency measurements only.

Since the sensitivity control is eliminated, AGC eases the job of an operator. A second advantage of AGC is its ability to handle input signals of time varying amplitude. AGC has limitations in measurement of high frequency signals with AM modulation. Since the AGC circuit makes adjustments for the measurement near the peak levels and ignores the valleys of the input signal, erroneous counting can result due to the presence of AM modulation in high frequency signals.

Time base oscillator considerations

The accuracy of the counter measurement depends upon the accuracy of the time base oscillator circuit used. Most counters employ a quartz crystal as the oscillating element. We will discuss the different types of crystal oscillators used in counters along with the basic factors that can affect accuracy of the oscillator.

Types of time base oscillators. The three basic types of crystal oscillators are:

- Room Temperature Crystal Oscillator (RTXO)
- Temperature Compensated Crystal Oscillator (TCXO)
- Oven Controlled Crystal Oscillator

RTXO. The Room Temperature Crystal Oscillators are manufactured for minimum frequency change over a range of temperature—typically between 0°C to 50°C. This is accomplished basically through the proper choice of the crystal cut during the manufacturing process. The frequency of a high quality RTXO would vary by about 2.5 ppm over the temperature range of 0°C to 50°C.

TCXO. The Temperature Compensated Crystal Oscillator uses compensating components to null the effect of change of crystal properties with temperature. Consider the electrical equivalent circuit of the quartz crystal as shown in Figure 9.10.

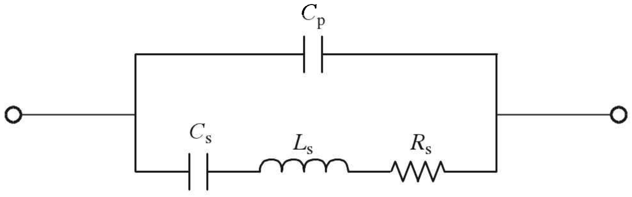


Figure 9.10 Equivalent circuit of a crystal.

The values of R_s, C_s, L_s, and C_p are determined by the physical properties of the crystal. An external variable capacitance is typically added to obtain a tuned circuit for oscillations. The temperature dependence of L, C and R make the frequency of the crystal oscillator temperature sensitive. To compensate for frequency changes due to temperature variation, an external capacitance or components with opposite temperature coefficient are connected to the crystal to obtain a more stable frequency of the tuned circuit. Oscillators with this method of compensation are often called Temperature Compensated Crystal Oscillators (TCXO). These oscillators offer an order of magnitude improvement in frequency stability over that of the room temperature uncompensated type. Typical frequency changes are 5×10^{-7} over 0°C to 50°C temperature range, or five times better than those of the RTXO.

Oven controlled crystal oscillator. The basic principle of this oscillator is to control the surrounding temperature of the crystal. The crystal oscillator is housed in an oven which minimizes the temperature changes surrounding the crystal. Two types of ovens are typically employed—the simple ON/OFF switching oven and the proportional oven. The simple switching oven turns the power OFF when the maximum temperature is reached and turns the power ON when the minimum temperature is reached. The more sophisticated proportional oven controls and provides a heating that is proportional to the differential between the actual temperature and the desired temperature surrounding the crystal oscillator inside the oven. Typical variation in frequency for a high quality proportional oven controlled crystal oscillator is less than 7 parts in 10^9 over the 0°C to 50°C temperature range.

It usually takes 24 hours or more after turn-ON for the oven oscillator to achieve its specified stability. However, it can come to 5 parts in 10^9 of the final specified frequency value after a 20 minute warm-up. Most counters employing an oven oscillator have a feature whereby the oscillator is powered whenever the power line is connected even if the counter is not turned on. Keeping the counter connected to the power line avoids the need for the warm-up phase and retrace.

Factors affecting accuracy of crystal oscillators. Apart from the temperature effects, the factors like line voltage variation, aging or long-term stability, short-term stability, magnetic fields, gravitational fields, and environmental factors such as vibration, humidity and shock also affect the accuracy of the crystal oscillators. The first three factors are the significant ones and are discussed next:

Effect of line voltage variations. Though all the counters use voltage regulators, the effectiveness of these regulators decides the changes in the oscillator frequency with the variations in the line voltage. Changes in the level of regulated voltage cause changes on bias levels and phase of feedback signal resulting in variation in the output oscillator frequency. A high stability, oven-controlled oscillator provides frequency stability of the order of 1 part per 10^{10} for 10% change in line voltage applied to the oven. For RTXO, the frequency stability is typically of the order of 1 part per 10^7 for the same 10% change in line voltage. Regulation better than this is unnecessary as frequency variations due to temperature effects would mask the effects of line voltage changes.

Aging rate or long-term stability. The physical properties of the quartz crystal exhibit a slow change with time resulting in a gradual cumulative frequency drift called aging. The aging rate is dependent on the inherent quality of the crystals used. Aging is often specified in terms of frequency changes per month since temperature and other effects mask the small amount of aging for a shorter time period. For a good RTXO, the aging rate is typically of the order of 3 parts per 10^7 per month. For a high quality oven controlled oscillator, the aging rate is typically 1.5 parts per 10^8 per month.

Short-term stability. Often referred to as the time domain stability, or fractional frequency deviation, short-term stability is the result of the inevitable noise (random frequency and phase fluctuations) generated in the oscillator, voltage transients, shock and vibration, cycling of crystal oscillator, etc. Since this noise is spectrally related, any specification of short-term stability must include the averaging or measurement time involved. The effect of this noise usually varies inversely with measurement time. Therefore, measurements are taken for longer gate times and by performing multiple period average measurements. With quoted averaging time, the specification of short-term stability essentially specifies the uncertainty due to noise in the oscillator frequency over the quoted time period (rms of the frequency variations given for different averaging times).

Table 9.1 summarizes the characteristics of different types of oscillators.

Table 9.1 Typical specifications of the four types of oscillators

Factor	*Room temperature crystal oscillator*	*Temperature compensated crystal oscillator*	*Simple switching oven oscillator*	*Proportional oven oscillator*
Temperature (0°–50°C)	$< 2.5 \times 10^{-6}$	$< 5.0 \times 10^{-7}$	$< 1.0 \times 10^{-7}$	$< 7.0 \times 10^{-9}$
Line voltage (10% change)	$< 1.0 \times 10^{-7}$	$< 5.0 \times 10^{-8}$	$< 1.0 \times 10^{-9}$	$< 1.0 \times 10^{-10}$
Long-term stability	$< 3.0 \times 10^{-7}$ per month	$< 1.0 \times 10^{-7}$ per month	$< 1.0 \times 10^{-7}$ per month	$< 1.5 \times 10^{-8}$ per month or $< 5.0 \times 10^{-10}$ per day
Short-term stability (1 sec average)	$< 2.0 \times 10^{-9}$ rms	$< 1.0 \times 10^{-9}$ rms	$< 5.0 \times 10^{-10}$ rms	$< 1.0 \times 10^{-11}$ rms

The total time base oscillator error or inherent error is the cumulative effect of all the individual sources of error described above. The time base error is only one of the several sources of measurement error for the counter. Hence, it may or may not be significant for a given counter measurement depending on the particular application involved. Sources of counter measurement errors are described in the following section.

Main gate considerations

The main gate of the counter exhibits propagation delays and takes some finite time to both switch ON and OFF. This finite amount of switching time is reflected in the total amount of time the gate is open for counting. If this switching time is significant compared to the period of the highest frequency counted, errors result in the count. However, if this switching time is significantly less compared to the period of the highest frequency counted, the error is not appreciable. For example, for a 500 MHz signal (2 ns period), this error is insignificant if the switching time of the main gate is substantially less than 1 ns. For true 500 MHz operation, high-speed devices are necessary in the gate, input and counting register circuitry.

9.2.3 Sources of Measurement Errors

Types of measurement errors

The major sources of measurement errors for an electronic counter are generally classified into the following four categories:

- ±1 count error
- Time base error
- Trigger error
- Systematic error

±1 Count error. When an electronic counter makes a measurement, a ±1 count error or ±1 count ambiguity or gating error can exist in the least significant digit. This is often referred to as *quantization error*. This ambiguity can occur because of the synchronization problem between the internal clock frequency and the input signal as illustrated in Figure 9.11. The

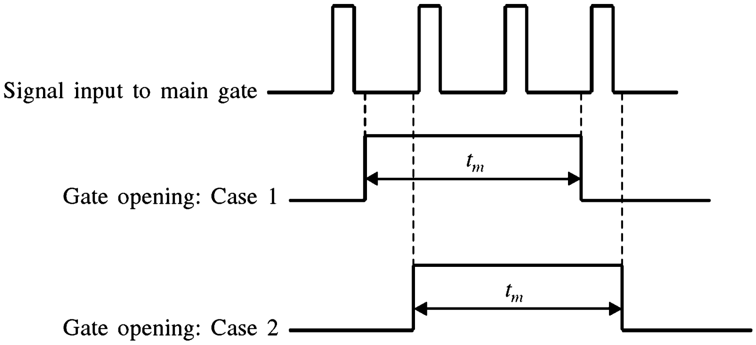


Figure 9.11 ±1 count ambiguity.

main gate is open for the same time t_m in both cases. Lack of synchronization between the clock and the input signal can cause valid counts to be skipped, which for this example are 2 for Case 1 and 3 for Case 2.

The error caused by this ambiguity is, in absolute terms, ±1 count out of the total accumulated count. For example, for input frequency of 10,000 Hz, the count will lie between 9,999 and 10,001 Hz, i.e. 10 kHz ±1 Hz. And if input frequency is 10 Hz, the count will lie between 9 and 11 Hz, i.e. 10 Hz, ±1 Hz.

We can define the relative frequency measurement error due to ±1 count ambiguity as

$$\frac{\Delta f}{f} = \frac{\pm 1}{f_A}$$

where f_A is the input signal frequency.

Therefore, the relative or % error is less for high frequency compared to low frequency input signal. To keep this error small, period measurement is preferred at low frequencies and frequency measurement at high frequencies.

Time base error. Any error resulting from the difference between the actual time base oscillator frequency and its nominal frequency is directly translated into a measurement error. This difference is the cumulative effect of all the individual time base oscillator errors described previously and is expressed as ppm (parts per million).

Trigger error. Trigger error is a random error caused by noise on the input signal and noise from the input channels of the counter. In period and time interval measurements, the input signal(s) control the opening and closing of the counter's gate. The effect of the noise is to cause one limit of the hysteresis window to be crossed too soon or too late—causing the main gate to be open for an incorrect period of time. This results in a random timing error for period and time interval measurements.

Systematic error. For time interval measurements, any slight mismatch between the rise times and propagation delays of the start channel and the stop channel amplifiers results in internal systematic errors. Mismatched probes or cable lengths introduce external systematic errors.

Trigger level timing error is another systematic error which is caused by uncertainty in the actual trigger point especially in the time interval measurements. This uncertainty is not due to noise, however, but is due to offsets in trigger level readings caused by hysteresis and drifts. Trigger level timing error may be expressed as:

$$\Delta T = \frac{\text{trigger level error}}{\text{signal slew rate at trigger point}}$$

Not all these four categories of measurement errors are significant for all modes of counter measurement. For example, only the ±1 count and time base errors are considered as important for frequency measurements using conventional counters. Table 9.2 summarizes the significant errors in various measurements.

In period measurement, all of the first three types of errors can affect the accuracy of the measurement, while all the four types of errors can be significant for time interval measurements.

Table 9.2 Summary of measurement errors

Source of errors	*Frequency measurement*	*Period measurement*	*Time interval measurement*
±1 count	Yes	Yes	Yes
±Time base	Yes	Yes	Yes
±Trigger		Yes	Yes
±Systematic			Yes

Frequency measurement error

The total frequency measurement error is defined as the sum of its ±1 count error and its total time base error.

Example 9.1

The frequency of 10 MHz is to be measured by the counter. Find the dominant error given the total error of the time base as 1 ppm, i.e. 1×10^{-6}.

Solution

The error contributed by the time base in the measurement of a 10 MHz signal is $1 \times 10^{-6} \times 10^{7}$ Hz or 10 Hz. Or, the relative frequency measurement error due to the time base error is $\pm\ 1 \times 10^{-6}$. The error due to the ±1 count error is $\pm 1/10^{7}$ or $\pm\ 1 \times 10^{-7}$ for a one second gate. Therefore, in this case, the time base error is dominant over the ±1 count error. (But the ±1 count error becomes dominant for input frequency less than 1 MHz).

Period measurement error

The total period measurement error may be defined as the sum of its ±1 count error, time base error and trigger error.

For period measurement, the signal counted is the internal time clock of frequency f_C and period t_C. Hence, the relative period measurement error due to ±1 count ambiguity is

$$\frac{\Delta T}{T} = \frac{\pm t_C}{T_A}$$

where T_A is the period of the input signal.

The relative period measurement error due to time base error is again the dimensionless factor expressed in ppm.

The general expression for computing the trigger error with single trigger in period measurement is:

$$\text{rms trigger error} = \frac{\sqrt{x^2 + e_n^2}}{\dfrac{\Delta V}{\Delta T}}$$

where x is the noise contributed by the counter's input channels (less than several hundred microvolts in some counters to as high as several millivolts in others), e_n is rms noise contributed by signal source measured over the counter's bandwidth and $\Delta V/\Delta T$ is the slew rate at trigger point of input signal.

Since for period measurements, trigger error occurs at the beginning and end of the measurement and the noise adds on an rms basis, the rms trigger error for period measurements (and frequency measurements made by a reciprocal counter) is then given by

$$\text{rms trigger error} = \frac{\sqrt{2}\sqrt{x^2 + e_n^2}}{\dfrac{\Delta V}{\Delta T}}$$

It is apparent from this expression that the trigger error can be reduced by input pulses with fast rise time or fast slew rate.

The ±1 count and the trigger error (but not the time base error) can be reduced by the multiple period averaging technique. The main gate is opened over several cycles of the input signal and the average period of the repetitive signal is determined. The multiple period averaging measurement error then can be written as

$$\pm\frac{1\,\text{count error}}{n} + \frac{\text{trigger error}}{n} \pm \text{time base error}$$

where n is the number of cycles that have been averaged.

It should be noted that the ±1 count in period (or period averaging) measurement refers to the counted clock while that for frequency measurement, the ±1 count is that of the input signal. The ±1 count and trigger errors are considered to occur randomly with a normal distribution and, hence, are reduced inversely as the number of cycles averaged is increased. The time base error factor (which is solely due to the total error of the time base) is not reduced by the period averaging technique. It should be noted, however, that the absolute magnitude of the time base error is dependent on the magnitude of the period being measured.

Example 9.2

The period of 100 ms and 1 s is to be measured by a counter with the time base of 1 ppm. Calculate the absolute time base error without and with averaging of 100 cycles in each case.

Solution

100 ms period:

Absolute time base error without averaging = $1 \times 10^{-6} \times 100$ ms = 100 ns

$$\text{Absolute time base error with averaging} = (1 \times 10^{-6})\,\frac{100 \times 1000}{1000}\ \text{ms} = 100\ \text{ns}$$

Therefore, averaging does not reduce the time base error.

1 s period:

$$\text{Absolute time base error without averaging} = 1 \times 10^{-6} \times 1\ \text{s} = 1\ \mu\text{s}$$

$$\text{Absolute time base error with averaging} = (1 \times 10^{-6})\frac{1 \times 1000}{1000}\,\text{s} = 1\ \mu\text{s}$$

Therefore, averaging does not reduce the time base error but the magnitude of the error increases by increasing the period from 100 ms to 1 s.

Time interval measurement error

The accuracy statement of the time interval (T.I.) measurement error may be written as

T.I. measurement error = ±1 count ± trigger error ± time base error ± systematic error

The ±1 count error in time interval measurement refers to one count of the clock frequency. Hence the higher the clock frequency, the smaller the ±1 count error.

The general expression for computing the trigger error for time interval measurement is given by

$$\text{rms trigger error} = \sqrt{\frac{x^2 + e_{nA}^2}{\left(\frac{\Delta V}{\Delta T}\right)_A^2} + \frac{x^2 + e_{nB}^2}{\left(\frac{\Delta V}{\Delta T}\right)_B^2}}$$

where x is the noise contributed by the counter's input, $e_{nA/B}$ is rms noise from source driving the A(START)/B(STOP) channel and $(\Delta V/\Delta T)_{A/B}$ is the slew rate at trigger point of input signal at A/B. It is apparent from this expression that the trigger error can be reduced by short input pulses with fast rise time or fast slew rate.

The time interval measurement error due to time base error is again the dimensionless factor expressed in ppm.

Systematic error is a fixed error and is repeated in every measurement. This is usually small but is important in absolute measurements of pulse width or time delays of short duration. Since the error is fixed, it can reduce the accuracy of the measurement but has no effect on the resolution.

The accuracy of the time interval measurement error can be improved in several ways. ±1 count and ± trigger errors are of a random nature and can be reduced by taking the statistical average of a large number of measurements. For a time interval averaging with N intervals averaged, these two sources of random error are reduced by a factor of $1/\sqrt{N}$. The reason for the square root is due to the fact that the random error can occur in all the start/stop gate operations required for each of the time interval measurements averaged.

The magnitude of the time base error can be reduced by the use of a better quality time base oscillator. The systematic error can be made insignificant through proper calibration of the measurement setup and elimination of the mismatch between the start and stop channels.

9.3 RECIPROCAL ELECTRONIC COUNTERS

The reciprocal electronic counter is a new class of counter which always makes a period measurement on the input signal. If frequency information is desired, it can be directly displayed by taking the reciprocal of the period measurement. The reciprocal technique offers two major features:

- The ±1 count quantization error is independent of the input signal frequency. Hence, for a noiseless input signal and assuming negligible trigger and time base error, the resolution of the reciprocal counter is also independent of the input signal frequency.
- The period counting characteristic of the reciprocal technique provides the capability for control of the main gate in real time, i.e. time base can be synchronized with the input signal. This ties down the resolution of the measurement with the time base and not the input signal. The result is the speeding up of the measurement of the low frequency.

We know that

$$\text{Relative frequency measurement error due to } \pm 1 \text{ count} = \frac{\pm 1}{f_A}$$

$$\text{Relative period measurement error due to } \pm 1 \text{ count} = \frac{\pm t_C}{T_A}$$

where f_A is the frequency of the input signal, t_C is the period of the clock, T_A is the period of the input signal or the gate time of the counter if the gate remains open longer than one cycle of the input signal.

We see that, for a given gate time, the amount of quantization error for frequency measurement is inversely proportional to f_A, the input frequency. In period measurement, for the same gate time, the quantization error is constant and is determined by t_C. The difference in quantization error of the two methods of measurement is shown in Figure 9.12.

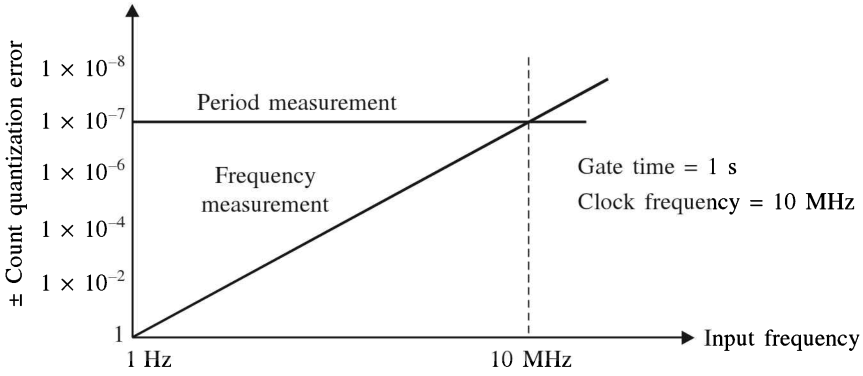


Figure 9.12 Difference in quantization error for period and frequency measurements.

As shown in Figure 9.12, for all input frequencies less than the clock frequency for a given measurement time, the ±1 count quantization error in the period measurement is always smaller than that for a corresponding frequency measurement. Assuming negligible trigger and time base errors, the period measurement always has a higher resolution (less ±1 count quantization error) than that of a corresponding frequency measurement for all input frequencies less than that of the clock frequency. It means that the reciprocal technique can achieve the same resolving capability of the conventional frequency measurement approach with a significantly less measurement time.

However, the ±1 count quantization error for the period measurement is larger than that for a corresponding frequency measurement for input frequencies higher than that of the clock frequency. In a modern reciprocal counter, the measurement mode is automatically switched over to the frequency mode for input frequency higher than the clock frequency. In this way, the counter achieves improved resolution for all admissible input frequencies.

Example 9.3

(a) For a 10 Hz signal with a 1 s gate and using a 10 MHz clock, the frequency measurement error is equal to $\pm 1 \times 10^{-1}$ and the period measurement error is equal to $\pm 1 \times 10^{-7}$ for 1 s gate time. What is the frequency shown by the reciprocal counter?

(b) What is the frequency shown by the basic frequency counter and the reciprocal counter, with a 1 s gate and using a 10 MHz clock, if the input signal frequency is 10.0001 Hz?

Solution

(a) The reciprocal counter will show frequency of 10 Hz as 0.000001 Hz, i.e. 10×10^{-7}.

(b) When the input frequency is equal to 10.0001 Hz, the basic frequency counter would contain a count of 10 for 1 s gate time, by definition, and will show 10 Hz frequency. The reciprocal period counter on the other hand would contain a count of 10 but time would be measured as (10/10.001) = 0.9999900 s, giving the frequency as (10/ 0.9999900) = 10.0001 Hz.

Note that in the reciprocal counter, the measurement gate does not have to be open for exactly 1 s, so a (1/2) s gate time is also possible. This increases the speed of measurement and is useful for pulsed and changing signals.

Interpolation

The resolution of the counter can further be increased by using interpolators. This is shown in Figure 9.13. The edges of the input signal and the clock rarely line up. Interpolators are used to measure time between the starting and stoping edges, shown as T_1 and T_2.

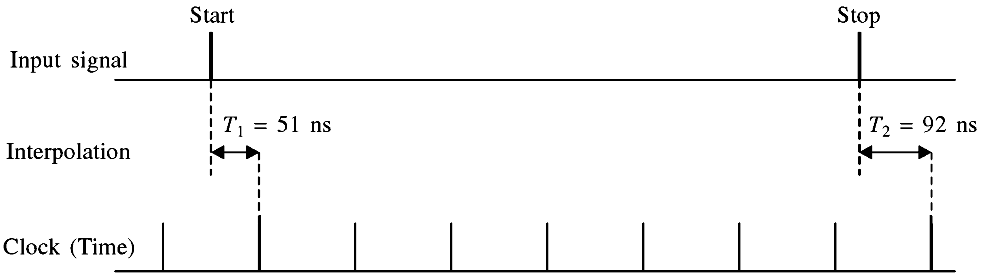


Figure 9.13 Time interpolation.

In the example shown in Figure 9.13, assume that the basic time base frequency is 10 MHz so each time base tic represents 100 ns. Let us assume that interpolator circuits can resolve time of 1 ns. The time measured by the counter between Start and Stop will be

$$T = 7 \times 100 \text{ ns} + T_1 - T_2 = 659 \text{ ns}$$

Therefore, the counter can measure time as 659 ± 1 ns, instead of 700 ns ± 100 ns which a counter without interpolators will measure. This gives the equivalent of 1 GHz time base or nine digits of frequency resolution with a 1 s gate time.

Arming

The second characteristic of the reciprocal counter is called *arming* or the capability of precise main gate control in real time. The arming capability is due to the fact that in period measurement, the input signal controls the opening/closing of the main gate. In frequency counting, the gate is controlled by the signal from the time base oscillator and the operator has little, if any, control on when the gate opens; all the operator knows is that at some undetermined point in time, the gate will open and accumulate counts from the input signal. The gate then closes at a precise interval of time later and the counter displays the average frequency of the input signal over the time the gate was open.

Figure 9.14 shows an example of the external arming. The arming signal is fed to the external arming input of the counter and the counter is set to gate based on the external arm. Figure 9.14 shows that the gate is made to open at the leading edge of the input signal following positive edge of the arming signal and the gate is stopped at the leading edge of the input signal following negative edge of the arming signal. This ensures the real time operation of the gate because the precise time of opening and closing of the gate is known.

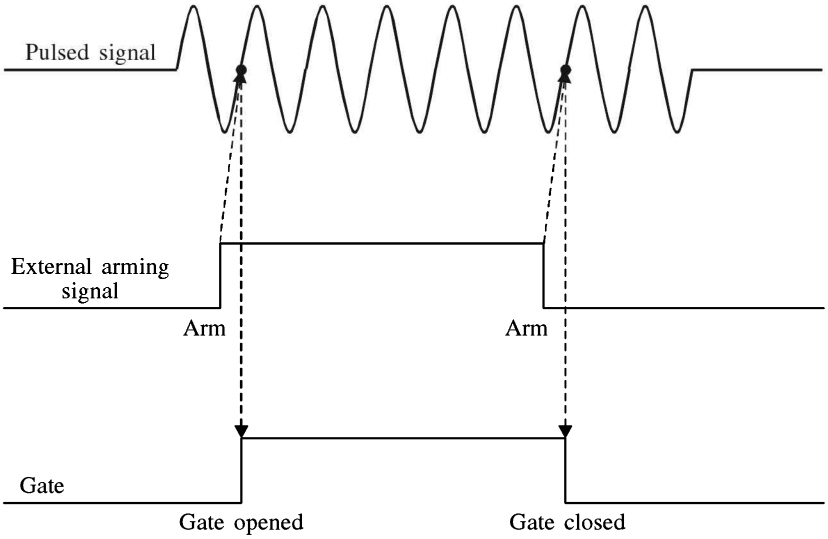


Figure 9.14 Relationship between input arming and gate signals.

Other techniques of operating the gate with the arming signal are available. Arming can be done automatically and many reciprocal counters offer only the automatic mode.

Block diagram of reciprocal counter

Figure 9.15 shows the block diagram of a reciprocal counter incorporating arming and interpolators.

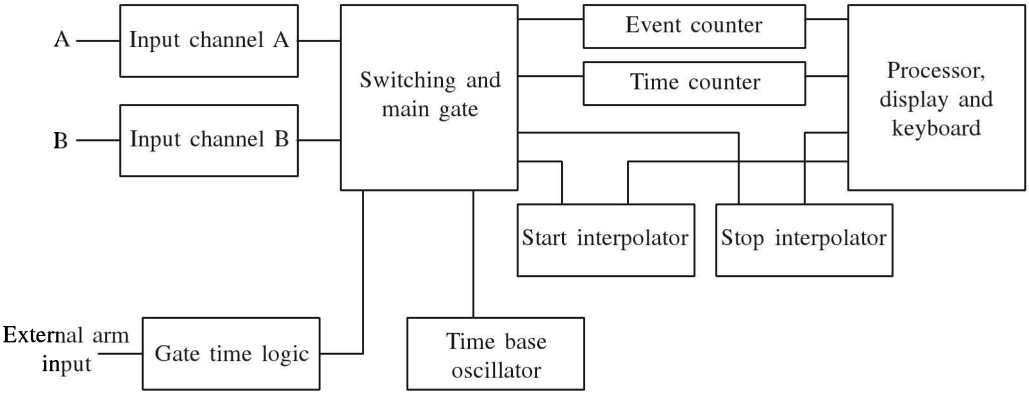


Figure 9.15 Block diagram of a reciprocal counter.

The basic block diagram of a reciprocal counter is essentially similar to the conventional counter except for the fact that the counting is done in separate registers for time and event counts. The contents of these registers are processed and their quotients computed to obtain either the desired period or frequency information which are displayed directly. The Event counter accumulates counts from the input signal while at the same time the Time counter accumulates counts from the internal clock for as long as the main gate is open. In a single period measurement, the main gate opens for precisely one period under the control of the input signal. During this time interval, the Event counter would have accumulated one count while the Time counter would have accumulated a number of clock pulses. The number of accumulated clock pulses is multiplied by the clock period to give the period of the input signal.

This computation is done automatically by the processor and the results are displayed directly. In period averaging, the main gate is open for more than one cycle of the input signals. The Event and Time counters accumulate and count pulses from the input signal and the internal clock, respectively, during this time while the gate is open. The quotient of the product of clock period and clock count to the event count is the average period of the input signal. In frequency averaging, the reciprocal of the quotient is automatically computed and the result is displayed as the average frequency.

Whenever external arming input is used, switching circuits detect the arming pulse and wait for the programmed point of the input waveform to start the gate. A similar operation takes place at the end of the arming pulse.

Time recorded by the Start and Stop interpolators is used by the processor to compute exact time of start and stop instants of the gate.

Features of reciprocal counter

The features of a reciprocal counter are summarized below:

- Utilizes the full resolving capability of the counter over its entire frequency range.
- The real-time measurement capability of period counting allows measurements on pulsed RF systems.
- Allows accurate low frequency measurements (e.g. power line frequency)
- High accuracy can be obtained in conveniently short measurement times.
- Use of microprocessor and other digital technology provide:
  - Greater arithmetical and computational capabilities such as statistics, offsetting or scaling made on the measurements.
  - Greater ease in using the instrument with self-check or calibration conveniences.
  - Improved interfacing and system capabilities.
  - Greater programmability.

9.4 SPECIFICATIONS OF ELECTRONIC COUNTERS

9.4.1 Input Characteristics

Range

It is the range of frequency over which the input amplifier's sensitivity is specified. If input coupling is selectable, then ac and dc are specified separately.

Example: dc COUPLED; dc to 100 MHz; ac COUPLED; 20 Hz to 100 MHz

Sensitivity

The lowest amplitude signal at a particular frequency which the counter can count. Assumes that the trigger level (if available) has been optimally set for a value equal to the midpoint of the input signal.

Sensitivity is actually a measure of the amount of hysteresis in the input comparator (Schmitt trigger) and may vary with frequency. Because of this, the sensitivity specification may be split into two or more frequency ranges.

The peak-to-peak minimum countable signal, defined as the counter's input amplifier sensitivity, is equal to UTL–LTL, where UTL is the upper-threshold voltage and LTL is the lower-threshold voltage. The rms sine wave sensitivity is given as

$$\text{rms sine wave sensitivity} = \frac{\text{UTL} - \text{LTL}}{2\sqrt{2}}$$

The counter input, normally, does not respond to the rms value of the waveform but only to the peak-to-peak value; the sensitivity specification should be therefore volts peak-to-peak with a minimum pulse width. Since many applications involve measuring the frequency of a sine wave signal, the specification is also given in terms of volts rms sine wave. Note, however, that a different waveform with the same rms voltage may not trigger the counter—the counter responds only to peak-to-peak.

Example: 10 mV rms sine wave to 10 MHz; 75 mV peak-to-peak pulse at minimum pulse width of 5 ns.

Signal operating range

This is the operating range of the input voltage. If the signal peaks extend beyond the specified signal operating range, one or more operating modes may give incorrect results; for example, frequency miscounting or time interval inaccuracies.

Example: –2.5 V to +1 V

Dynamic range

This is the minimum to maximum allowable peak-to-peak signal range, centred on the middle of the trigger level. The trigger level lies within the signal operating range. If the input signal exceeds this range, then the input amplifier may saturate, causing transitions of the input to be missed. The dynamic range is limited by the range over which the differential input of the amplifier can swing without saturation.

For some input amplifiers, the dynamic range puts a further restriction on the allowable signal peaks as specified by the signal operating range. The signal peaks must always stay within the signal operating range specification and the peak-to-peak value must stay within the maximum dynamic range specification.

Example: 100 mV to 1 V peak-to-peak pulse

Trigger level

Trigger level is usually the voltage at the centre of the hysteresis band and practically is the dc voltage applied to one input of the input comparator. Some counters have a readout of trigger level and indicate the settability and the accuracy of the trigger level. The settability specification indicates what tolerance trigger level may be set, and the accuracy specification indicates the worst case difference between the indicated trigger level and the actual trigger point.

9.4.2 Operating Mode Specifications

Range

This is the minimum value of the input which can be measured and displayed by the counter up to the maximum value of the input. The maximum time interval is what can be measured before the counter overflows.

Example: Frequency range: 0.1 Hz to 100 MHz; Time interval range: 100 ns to 10^5 s; Time interval average range: 0 ns to 10^5 s.

LSD displayed

The "LSD displayed" is the value of the rightmost or least significant digit in the counter's display. The LSD Displayed may vary with gate time and magnitude of the measured quantity. For example, the LSD Displayed by a reciprocal counter varies with gate time and frequency. Even for the case where the input quantity is perfectly stable, the counter's readings may fluctuate. This fluctuation is due to quantization error (±1 count error). The value of the LSD Displayed often is the same as the quantization error which represents the smallest non-zero change which may be observed on the display. Because of this, resolution and accuracy statements often specify quantization error as ± LSD Displayed.

Example: Frequency LSD Displayed:

Direct count:	0.1 Hz, 1 Hz, 10 Hz switch selectable
Prescaled:	10 Hz, 100 Hz, 1 kHz switch selectable

For direct count, the number of cycles of the input are totalized during the gate time. For a 1 s gate time, each count represents 1 Hz (the LSD Displayed). For a 10 s gate time, each count represents 0.1 Hz. For a prescaled input, the number of counts accumulated during the gate time is reduced by the prescale factor over what would be accumulated by a direct count counter. Consequently, for a 1 s gate, each count represents N Hz where N is the prescale division factor.

Example: Period LSD Displayed:

100 ns to 1 fs (femptosecond = 10^{-15} second) depending upon gate time and input signal.

Example: Time Interval LSD Displayed:

100 ns to 10 ps depending upon gate time and input signal.

Resolution

Resolution is the maximum deviation (or may be expressed as rms deviation if random noise is resolution limiting) between successive measurements under constant environmental and constant input conditions, and over periods of time short enough so that the time base aging is insignificant. Practically speaking, it represents the smallest change in the input quantity which can be observed in the measurement result. Random readings are due to quantization error and trigger error (for reciprocal counters).

Resolution can never be better than the LSD Displayed. Sometimes, the resolution is equal to the LSD Displayed. Often, the resolution is not as good as the LSD Displayed because of jitter and trigger errors which are random noise induced errors.

For counters using interpolators, the timing resolution may not always be a decade value like 1 ns but might be 2 ns. In such cases, because decade value gates are often used, the counter might show an LSD that is good to ± 2 counts or so.

Since resolution is often limited by noise (noise on the input signal or noise from the counter), it is sometimes specified as rms (1 s) deviation about the average value or may have rms terms as part of the specification.

If averaging is being used and the counter averages N measurements, it gains resolution of only by a factor of $\sqrt{N}$. So taking 100 measurements gains only one more digit of resolution, not two.

Example:

$$\text{Frequency resolution:} \quad \pm \text{LSD} \pm \sqrt{2} \times \frac{\text{trigger error}}{\text{gate time}} \times \text{frequency}$$

$$\text{Period resolution:} \quad \pm \text{LSD} \pm \sqrt{2} \times \frac{\text{trigger error}}{N}$$

where N equals the number of periods in the period average measurement.

$$\text{Time interval average resolution: } \pm \text{LSD} \pm \frac{\text{start trigger error}}{\sqrt{N}} \pm \frac{\text{stop trigger error}}{\sqrt{N}}$$

where N = gate time × frequency.

Example 9.4

Let the input channel noise be 250 μV and the input signal be 1 mV rms noise over the bandwidth of the front-end of the computer. The counter is measuring a 1 MHz sine wave with a 1 s gate time. The ± 1 count error is 0.001 Hz. The input rises to 350 mV in 1 μs. Find the displayed frequency resolution.

Solution

$$\text{LSD} = 0.001 \text{ Hz}$$

$$\text{Trigger error} = \frac{\sqrt{(250)^2 + (1000)^2}\,\mu\text{V}}{350\,\text{mV}/1\,\mu\text{s}} = 3 \times 10^{-9}$$

$$\begin{aligned}\text{Frequency resolution} &= \pm \text{LSD} \pm \sqrt{2} \times \frac{\text{trigger error}}{\text{gate time}} \times \text{frequency} \\ &= \pm\, 0.001 \pm 1.4 \times 3 \times 10^{-9} \times 1 \text{ MHz} \\ &= \pm\, 0.005 \text{ Hz}\end{aligned}$$

This means that the display's least significant digits may change by about ± 0.005 counts.

Example 9.5

Calculate the time interval (TI) measurement resolution of the signal shown in Figure 9.16. Let the input channel noise be 80 μV. The input frequency is 500 kHz and the gate time is 1 s. The ±1 count error is 100 ps.

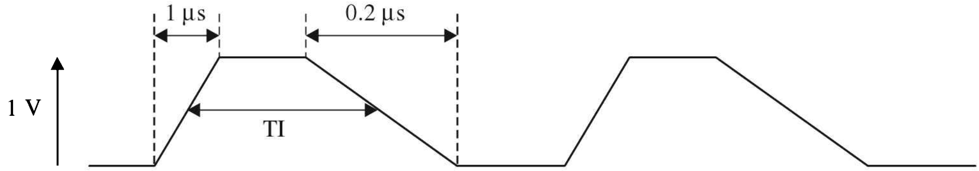


Figure 9.16 Waveform for calculating TI measurement resolution.

Solution

$$\text{Start trigger error} = \frac{80\ \mu\text{V}}{1\ \text{V}/\mu\text{s}} = 80 \times 10^{-12}\ \text{s}$$

$$\text{Stop trigger error} = \frac{80\ \mu\text{V}}{1\ \text{V}/0.2\ \mu\text{s}} = 16 \times 10^{-12}\ \text{s}$$

$$\text{TI measurement resolution} = \pm 100\text{ps} \pm \frac{80\ \text{ps}}{\sqrt{50 \times 10^4}} \pm \frac{16\ \text{ps}}{\sqrt{50 \times 10^4}}$$

$$= 100.137\ \text{ps}$$

Accuracy

Accuracy is defined as the closeness of a measurement to the true value as fixed by a universally accepted standard. The measure of accuracy, however, is in terms deviation from true value, or limit of error. High accuracy has a low deviation and low accuracy a high deviation.

Except for time interval measurements, accuracy is specified as resolution plus time-base error. Time-base error is the maximum fractional frequency change in the time-base frequency due to all error sources (e.g. aging, temperature, line voltage). The actual error in a particular measurement is found by multiplying the total $\Delta f/f$ of the time-base (or $\Delta t/t$ which equals $\Delta f/f$) by the measured quantity.

Accuracy = resolution ± (time base error × measured quantity)

For time-interval measurements, two additional errors are present: differential channel delay and trigger level timing error. Differential channel delay is the difference in delay between the start channel and the stop channel.

Trigger level timing error is the time error (in seconds) due to error in the trigger level readout [see Figures 9.17(a) and (b)]. If there is no readout of trigger level, then this error cannot be specified.

Unlike trigger error, trigger level timing error is a systematic error. The major component of the trigger level error comes from the fact that the actual trigger level is not a measurable level. The only dc level which can be measured is the output generated by the trigger level control which sets the level on one side of the comparator (Schmitt trigger circuit). This voltage represents the centre of the hysteresis band.

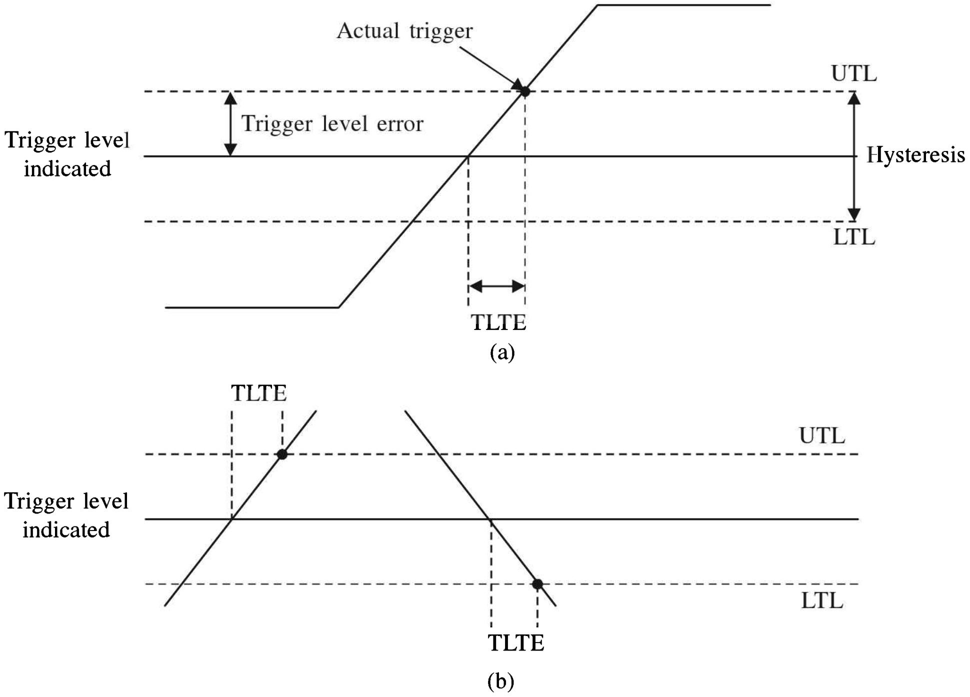


Figure 9.17 Trigger level timing error (TLTE).

In the absence of a specification, the trigger level error may be estimated at ±1/2 hysteresis band. The hysteresis band is equal to the peak-to-peak pulse sensitivity.

Figure 9.18 summarizes the components of the accuracy.

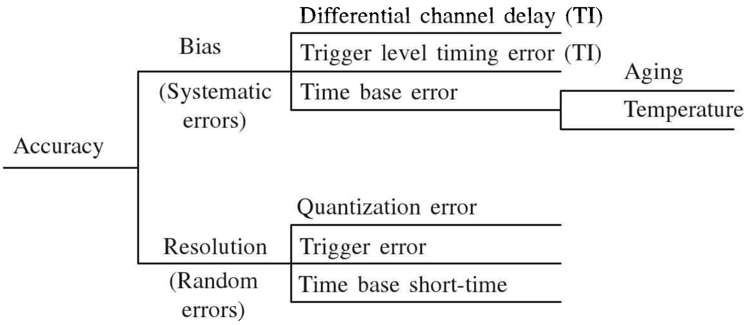


Figure 9.18 Components of the accuracy.

Example 9.6

Using the value of frequency resolution as calculated in Example 9.3 and the following data, calculate the frequency accuracy. Assume that counter has not been calibrated for one year.

Aging rate < 3×10^{-7}, Temperature effect < 5×10^{-6} (0 to 50°C), and Line voltage effect < 1×10^{-7} (±10% variation).

Solution

The total fractional frequency change in oscillator frequency is the sum of all effects.

$$\text{Total } \frac{\Delta f}{f} \text{ due to aging} = 3 \times 10^{-7} \times 12 = 3.6 \times 10^{-6}$$

For worst case temperature variations,

$$\text{Total } \frac{\Delta f}{f} \text{ due to temperature} = 5 \times 10^{-6}$$

For worst case line voltage variations,

$$\text{Total } \frac{\Delta f}{f} \text{ due to line voltage} = 1 \times 10^{-7}$$

$$\text{Total } \frac{\Delta f}{f} \text{ due to all sources} = 8.7 \times 10^{-6}$$

The worst case frequency change for 1 MHz would be $8.7 \times 10^{-6} \times 10^{6} = 8.7$ Hz
Therefore, frequency accuracy = ±0.005 Hz ± 8.7 Hz = ± 8.705 Hz

QUESTIONS

9.1 What is a universal counter? List the measurements which can be performed using a universal counter.

9.2 Draw the block diagram of a conventional frequency counter. Explain the operation of each block.

9.3 (a) Explain how is the input signal conditioned in a conventional frequency counter.
(b) Explain a typical time base circuit used in a conventional frequency counter.

9.4 Explain normalizing, presetting and prescaling functions used in a conventional frequency counter.

9.5 Draw the block diagram of a prescaled conventional frequency counter. Explain the operation of each block.

9.6 Draw the block diagram of a conventional frequency counter used for the measurement of *period*. Explain the operation of each block.

9.7 (a) What is meant by multiple period averaging technique?
(b) When is period measurement preferred over frequency measurement?

9.8 Draw the block diagram of a conventional frequency counter used for the frequency ratio measurement. Explain the operation of each block.

9.9 Explain the following parameters of the input conditioning circuit used in a conventional frequency counter:

- Sensitivity
- Input coupling
- Input impedance
- Dynamic range

9.10 Explain the following parameters of the input conditioning circuit used in a conventional frequency counter:

- Trigger level
- Trigger slope
- Automatic gain control

9.11 Explain the types of crystal oscillators used in time base oscillators? What factors affect the accuracy of crystal oscillators?

9.12 Explain the following errors in a conventional counter:

- ± 1 count error
- Time base error
- Trigger error
- Systematic error

9.13 (a) Define the total frequency measurement error in terms of the errors in conventional counters.

(b) In a frequency measurement, for what input frequency, ± 1 count error will dominate over the total time base error, given that the total time base error is 10 ppm.

9.14 The frequency of 100 MHz is to be measured by a frequency counter. Find the dominant error, given that the total error of the time base is 10 ppm.

9.15 (a) Define total period measurement error in terms of the errors in conventional counters.

(b) Show that time base error factor in not reduced by period averaging technique.

9.16 Define total time interval measurement error in terms of the errors in conventional counters. How can the accuracy of time interval measurement be improved?

9.17 Draw the block diagram of a reciprocal electronic counter. Explain the operation of each block. What are its major features?

9.18 Using typical values, plot ± 1 count quantization error versus input frequency in the case of period and frequency measurements using a reciprocal electronic counter. Explain the plot.

9.19 Explain the terms interpolation and arming. Draw the block diagram of a reciprocal electronic counter incorporating interpolation and arming.

9.20 Define resolution of electronic counters. How do you calculate frequency resolution, period resolution and time interval average resolution?

9.21 Let the input channel noise be 250 μV where the input signal has 1 mV rms noise over the bandwidth of the front-end of the computer. The counter measures a 2 MHz sine wave with a 1 s gate time. The ± 1 count error is 0.001 Hz. The input rises to 350 mV in 1 μs. Find the displayed frequency resolution.

9.22 Calculate the time interval (TI) measurement resolution of the signal shown in Figure Q9.22. Let the input channel noise be 80 μV. The input frequency is 100 kHz and the gate time is 1 s. The ± 1 count error is 50 ps.

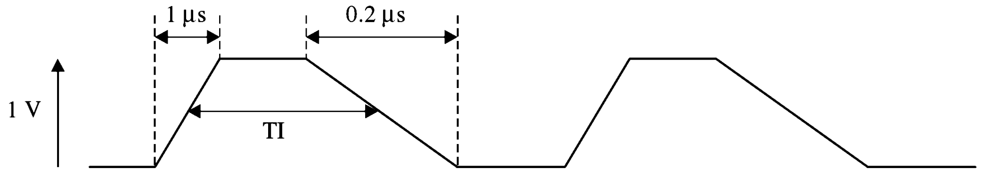


Figure Q9.22 Waveform for calculating the TI measurement resolution.

9.23 Define accuracy of electronic counters. How do you calculate frequency accuracy, period accuracy and time interval accuracy?

9.24 Define the various components of accuracy of electronic counters.

9.25 Use frequency resolution as calculated in Q9.21 and the following data and then calculate the frequency accuracy. Let us assume that the counter has not been calibrated for one year.

Aging rate < 3×10^{-7}, Temperature effect < 5×10^{-6} (0 to 50°C), and line voltage effect < 1×10^{-7} (± 10% variation).

9.26 Write the typical specifications of a universal counter.

SECTION II

Instrument Design and Manufacturing Techniques

Chapter 10

Grounding and Shielding

10.1 INTRODUCTION

Good design of electronic circuits and systems begins with grounding and shielding. Grounds and shields improve safety and reduce interference from noise. Properly grounded circuits and instruments reduce dangerous voltage differentials between them. Shields minimize interference from noise by reducing noise emissions and noise susceptibility.

The concept and procedure of grounding is extremely important in the design of electronic circuits and PCBs, installation of instrumentation panels, installation of analog and digital systems, etc. Generally, the users do not pay much attention to the grounding and shielding techniques, resulting in increase in the failure rate and electric shock hazards.

Various standards are available which give guidelines for safety while installing electric and electronic systems in houses and in plants. It is essential to follow these guidelines installing plant ground wiring to obviate any hazard.

We will now-study the concept and practices involved in effective grounding and shielding in electronic circuits and in industries.

10.2 GROUNDING

The term *ground* in electrical circuits refers to a current return path through the earth. Unfortunately, it has been loosely used to represent any type of current return path to an energy source. One of the first electrical symbols used is that shown in Figure 10.1.

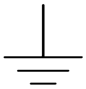


Figure 10.1 The symbol of earth ground.

This symbol represents a current return path through the earth to the low potential (voltage) side of an energy source. Frequently, however, it is used in electronic schematic drawings to indicate a current return such as a wire. In many cases it is used interchangeably with other symbols which, as we will see, are available to indicate specific returns. In any event, this use of the ground symbol can cause some confusion since many instruments provide an earth ground terminal.

10.2.1 The Concept of "Earth" Ground

Early developers of electrical systems theorized that the earth was an electrically neutral body, i.e. an equal number of negative and positive charges are distributed throughout the earth at any given time. Being electrically neutral, earth is considered to be at zero potential and establishes a convenient reference frame for voltage measurements. Noting that voltmeters read only the difference in potential between two points, absolute measurements can be made by using earth as a reference.

A true earth ground, as defined by the National Electrical Code, physically consists of a conductive pipe or rod driven into the earth to a minimum depth of 8 feet.

Figure 10.2 shows this concept, where the earth is used as the conductive current return path to the lowest potential point of the generating system

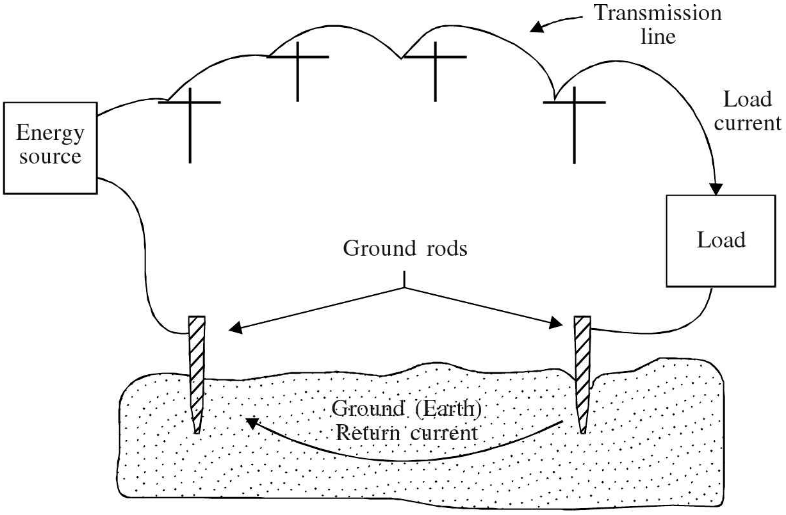


Figure 10.2 A transmission system using ground as a current return path.

10.2.2 An Illustration of a Typical Power Supply Grounding Error

As previously pointed out, the ground symbol, in many cases, has been taken as a generic symbol in electronic circuit diagrams to represent the current return path, even though no physical earth ground is used. This can cause some confusion when using instruments having an earth ground terminal. As an example, Figure 10.3 shows the front panel of a typical power supply. The supply is represented as a variable voltage battery. Note that three terminals are shown: a positive, a negative, and a ground terminal. The ground terminal of the supply is tied to the case of the instrument, which in turn is wired to a true earth ground.

Figure 10.3 depicts that using the positive terminal of the battery and the ground terminal does not complete a current return path to the energy source (battery), so no current will flow from the source, i.e. $I_{\text{load}} = 0$.

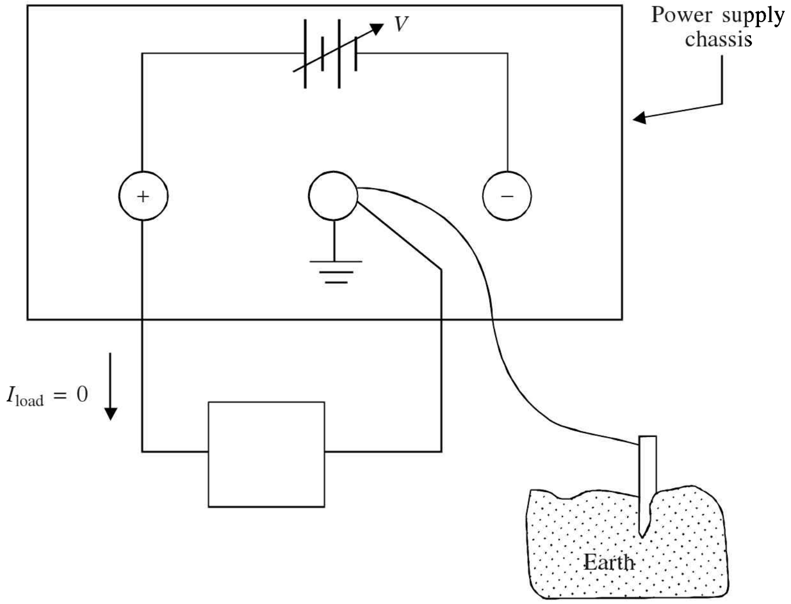


Figure 10.3 The connection of a power supply.

The positive and negative terminals must be used to complete a return path. Many circuits require both positive and negative voltages. A power supply that can provide each polarity must then be used (see Figure 10.4). The supply for positive voltage has the negative

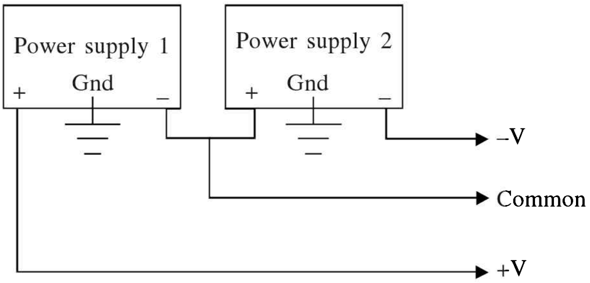


Figure 10.4 Power supply configuration for dual polarity voltages.

terminal as a return, and the negative supply has the positive terminal as the return. These two terminals are connected together, forming a common return path for load current. Figure 10.4 shows the proper connections for these supplies to provide both positive and negative voltages.

Although the common return path may be shown as a ground in the circuit diagram, the connection between the negative and positive terminals of the supplies results in a common or floating return.

10.2.3 Some Examples of Current Return Path Symbols

A current return path to an energy source is not necessarily, and frequently is not, earth ground. It can be a simple wire or a metal chassis or enclosure on which the circuit is mounted. Because the return is the point of lowest potential for all these cases, it is a convenient reference for circuit voltage measurements. Figures 10.5(a), 10.5(b), and 10.5(c) illustrate the symbols commonly used to represent the power supply common or signal or floating return, the chassis or signal ground, and the earth or safety ground returns, respectively. When more than one ground is required, the schematic circuit diagram generally defines the meaning of each symbol.

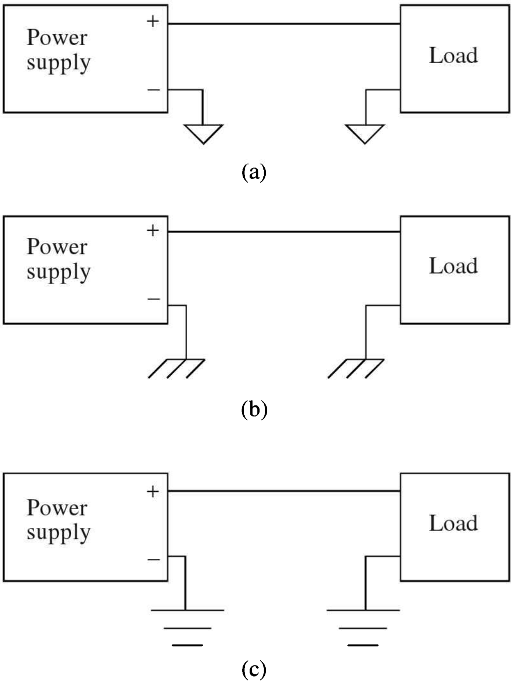


Figure 10.5(a) Common (floating) return, (b) chassis return, and (c) ground return.

10.2.4 Shock Hazard Protection Using Earth Ground

In instances where high voltages are required and chassis grounds or metal frames are used as return paths, hazardous conditions can be created if earth grounds are neglected. When the load circuit uses a metal enclosure as a chassis ground, resistive leakage or "sneak" paths can exist which result in high voltages between the enclosure and earth ground. Leakage is any unsuspected, unwanted resistive path between two points. If inadvertently, an earth-grounded object such as a water pipe, and the enclosure are simultaneously touched, a serious shock will result. Such a condition is illustrated in Figure 10.6.

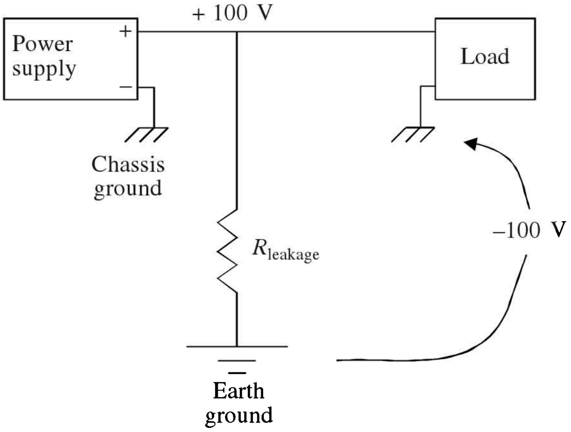


Figure 10.6 A shock hazard created by a leakage path.

In Figure 10.7, the earth ground is connected to the load enclosure, eliminating the shock hazard.

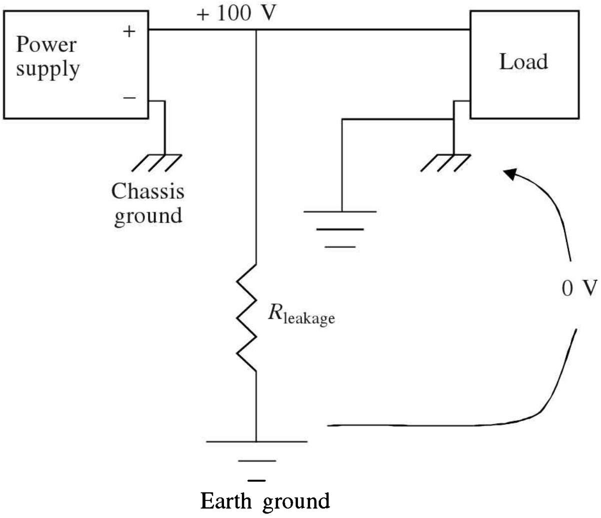


Figure 10.7 Using earth ground to eliminate a shock hazard.

10.2.5 Grounding Considerations

The most common noise problem encountered in large-scale electronic systems stems from a lack of good grounding practice. Grounding causes a lot of concerns to practising design and systems engineers. We will mention some basic practices to avoid grounding problems in the circuits.

If several points are used for ground connections, differences in potential between the points can cause troublesome "ground loops" which will cause errors in voltage readings. This is illustrated in Figure 10.8, where two separated chassis grounds are used. V_{gnd} represents a voltage existing between signal ground and the load ground. If voltage measurements are made between the load ground and the input signal, V_S, an erroneous voltage, $V_S + V_{\text{gnd}}$, will be indicated.

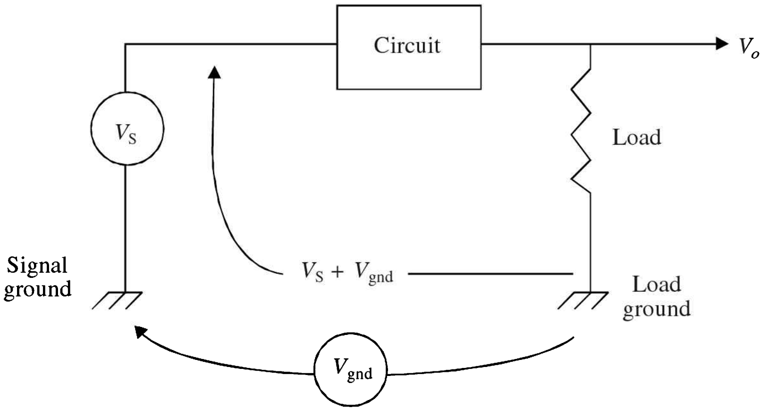


Figure 10.8 Generation of ground loop.

A common sign that a ground loop(s) exists, or that a ground is missing, is the presence of induced power line (50 Hz) noise in the circuit. Finding and eliminating troublesome ground loops in complex electronic systems can be a difficult and frustrating task; it requires an expertise gained largely through experience. This is why grounding, in many cases, is referred to as a *black magic art.*

10.2.6 Basic Grounding Practices

Circuit grounding

The ideal "single point ground (or return)" concept ensures that no ground loops are created. As the name implies, all circuit grounds are returned to a common point. This concept is shown in Figure 10.9.

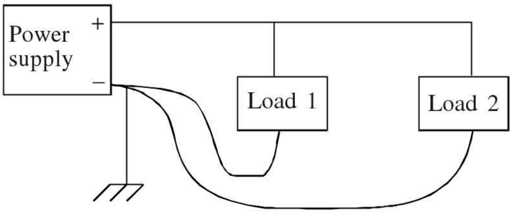


Figure 10.9 The concept of "single point ground".

If a circuit has many ground points, connecting them to a common point may be a difficult job. The next alternative is a ground bus.

Ground bus

Bus bars are available or can be constructed to serve as an adequate substitute for single point ground. The bus bar is simply a heavy wire or copper bar of low resistance which can carry the maximum sum total of the load current back to the power supply. The bus can be extended along the length of the circuitry so that convenient connections can be made at various points. The use of a ground bus is shown in Figure 10.10.

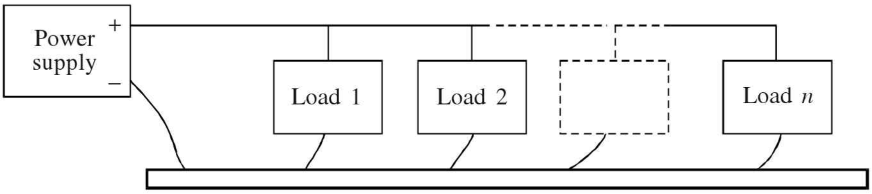


Figure 10.10 Bus bar return.

Ground (return) plane

A ground plane is very important component of any circuit. In essence, it is the return path for all signals including the power distribution. The ground plane can be thought of as homogeneous for the dc power only. In all other situations it is strictly inhomogeneous. This means that all grounds are not the same. As various circuits use the ground plane for their signal and power return paths, currents—conducted and induced—are caused to flow throughout the ground plane, and potentially can affect any or all other circuits, and can cause real problems. Multilayer PCBs generally use separate layers of supply and ground planes.

A changing current in the plane creates the changing magnetic flux which generates voltage as given by the expressions.

$$\Phi = BA = \mu_0\, niA$$

$$v = d\Phi/dt = A(dB/dt) = A\mu_0 n(di/dt)$$

where Φ is the magnetic flux, B magnetic field, A loop area, μ_0 permeability of free space, n number of turns in the loop, i the current, and v the voltage.

It is clear that to reduce the voltage, either the current or the area should be reduced. Since inductance is a function of loop area, reducing the loop area will reduce the inductive reactance of a circuit. The current follows the path of lowest impedance, not necessarily lowest resistance. Therefore the current will follow the path of minimum inductive reactance; this means that current flow will minimize the loop area in a circuit as shown in Figure 10.11.

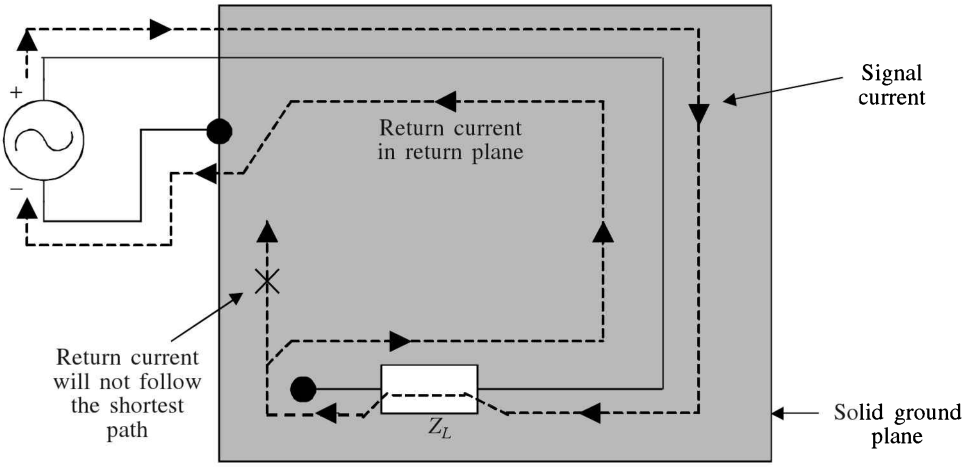

+
–
Signal current
Return current in return plane
Return current will not follow the shortest path
Z_L
Solid ground plane

Figure 10.11 Path of return current.

Note that a slot in the ground plane of a circuit board will increase the loop area of the circuit (see Figure 10.12). Such slots should be avoided because they increase the chance of noise coupling into other circuits.

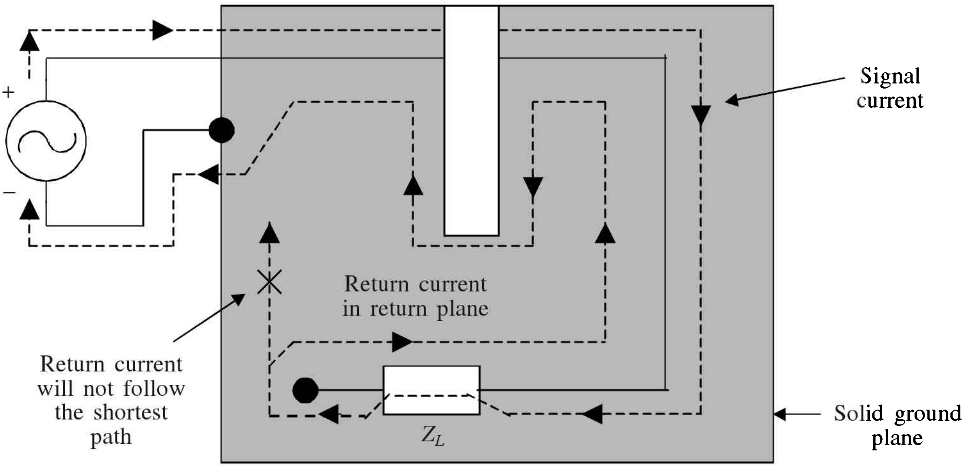

+
–
Signal current
Return current in return plane
Return current will not follow the shortest path
Z_L
Solid ground plane

Figure 10.12 A slot in return path.

10.2.7 Examples

Interfacing an ADC between analog and digital circuits

The analog-to-digital converter (ADC) is one application that needs a single-point ground for signal reference. Separate references can generate noisy ground loops. Figure 10.13 illustrates how a single-point reference eliminates large loops.

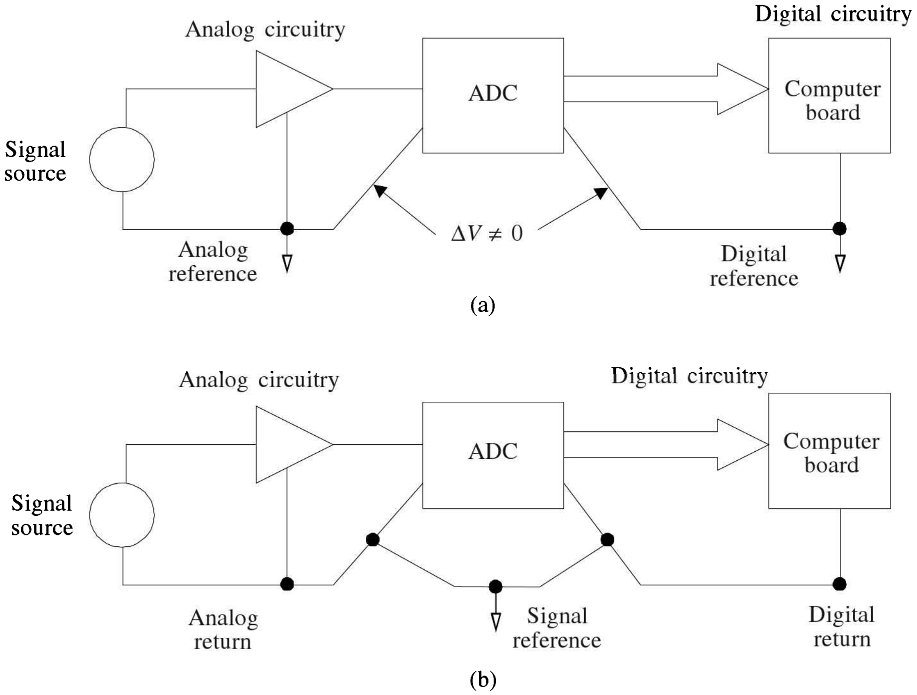


Figure 10.13 Interfacing an ADC between analog and digital circuits: (a) bad design and (b) better design.

Multi cabinet grounding

A large number of cabinets having analog and digital circuits are required in the industries. These must be properly grounded. Figure 10.14 shows a typical multicabinet grounding procedure. In the figure, AG denotes Analog Ground, DG stands for Digital Ground and CG for Cabinet Ground. Analog and digital circuits have separate power supplies. Within a cabinet, all power returns of analog circuits should be tied together to a single point AG. Similarly, all power returns of digital circuits should be tied together to a single point DG. Generally, thick insulated wires (4 AWG) are used for plant grounds.

The earth ground for all CG grounds should be within 1 Ω of true earth ground and within 1 Ω of each other at 25°C. If plant grounds are not within this limit, CG grounds are connected with 0000 AWG insulated ground cable instead of insulated 4 AWG wire.

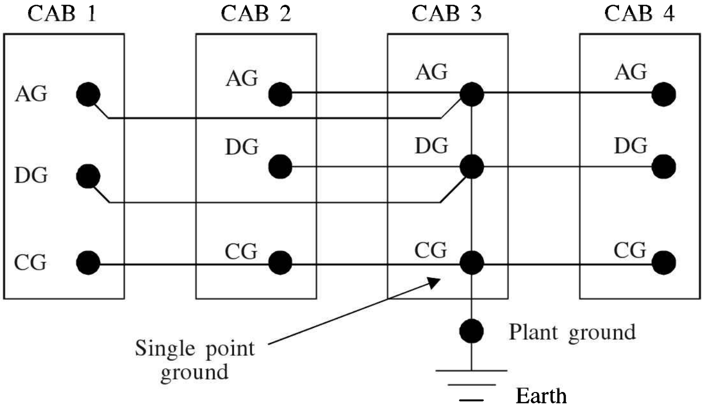


Figure 10.14 Multicabinet grounding.

Multilayer PCB

The use of multilayer printed circuit boards allows the use of multiple ground planes, as well as buried (under the signal layers) V_{CC} and ground layers. These layers are sandwiched together and act as a very efficient distributed bypass capacitor. A variation on this is to have the V_{CC} and ground layers as the outer or intermediate layers, thus shielding the buried signal layers; or some combination thereof.

Figure 10.15 shows an example where multiple ground planes are used in the multilayer boards. VIA is a hole through a PCB, connecting various unattached layers but serving as a

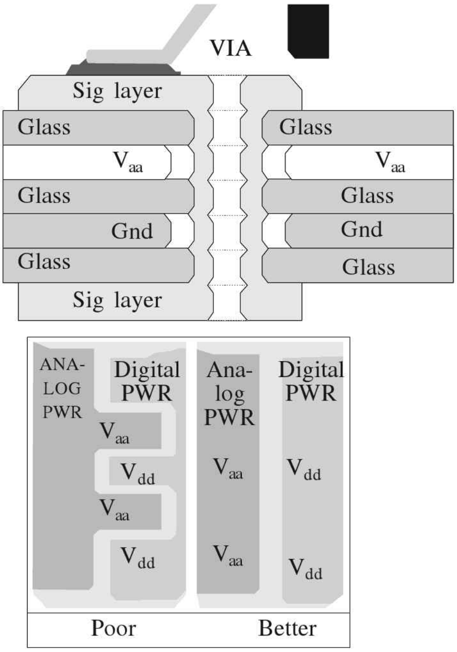
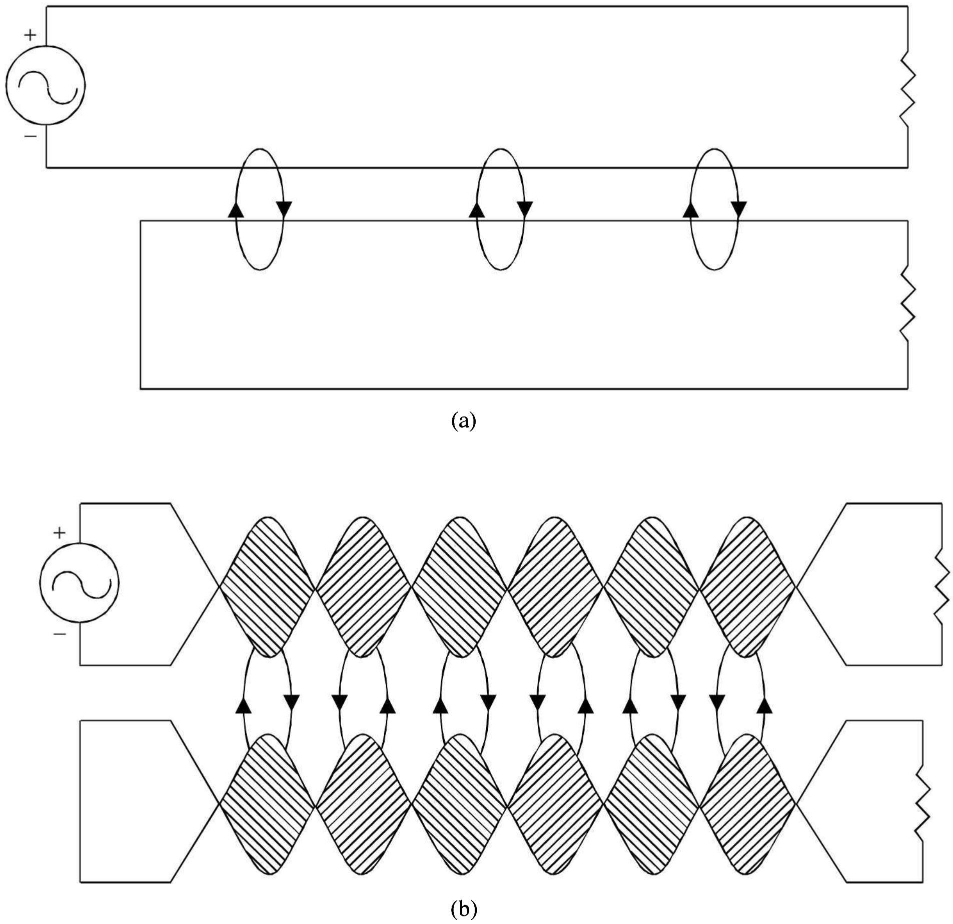


Figure 10.15 Ground and power supply planes in multilayer boards.

component anchor point. VIAs are usually plated. VIA can be a Blind VIA (that does not go completely through the board, but is instead attached to an internal layer) or a buried VIA (that does not show on either surface of the PCB, but runs between internal layers).

Use of twisted wires

One of the most common areas of inductive coupling is in cables that connect various circuits or boards. The long, straight wires encompass a significant loop area that provides an inductive reactance and, hence, inductive coupling. Twisting the pairs of signal and return lines together eliminates the loop area and the mutual inductive coupling between the circuits. (See Figure 10.16). To get proper inductive shielding, signal wire should always be paired with the return wire, i.e. they should form a pair, otherwise we will not gain any inductive shielding.



(a)

(b)

Figure 10.16 Effect of (a) straight wires and that of (b) twisted wires.

10.3 SHIELDING

Shielding either prevents noise energy from coupling between circuits or suppresses it. Shielding can be anything from using a coaxial or shielded cable, to a sealed conductive chamber for circuit isolation. Shielding serves a reciprocal purpose: it protects the circuit it is shielding from outside noise or unwanted signals; and conversely, it contains its own signals and thus protects the outside world from interference of its own making. Shielding is mostly used to block electrostatic or "E" fields (Faraday shield). However, if ferrous metal (tempered Mu Metal works best for magnetic fields) is used, then both electrostatic and some level of magnetic shielding is accomplished. This is especially useful where open frame transformers or unshielded coils are used and would otherwise exchange signals by mutual inductance.

One important requirement for a shield to be effective, is that there must be no currents flowing through the shield itself. This is best accomplished by connecting the reference or common, at only one point on the shield, thus preventing any flow of current. The reason for this is that, any current flowing in the shield material itself can produce secondary fields on the other side of the shielding material and thereby reducing the effectiveness of the shield. An extreme case of this might be a shielded cable, whose shield has a different potential at each end, and therefore the resulting current flow in the shield induces unwanted noise into the centre or shielded conductors. In this situation, we might find a remedy by disconnecting one or the other ends of the cable's shield. However, this may not prove satisfactory in certain environments, and may require a *guard* potential, or better still: optical isolation.

At high frequencies (>1 MHz), multiple ground connections are necessary because stray capacitance at the ungrounded end of a shield can complete a ground loop. Therefore, we should ground both ends of a long (relative to wavelength) shield.

A metal enclosure can be an effective electrostatic shield or Faraday cage. It prevents capacitive coupling and need not be grounded if it completely encloses a circuit. Grounding, however, ensures that current from stray capacitance flows to the signal reference ground rather than feeds back and causes cross talk.

Another source of noise, from which proper shield is required, is electromagnetic interference (EMI) from lightning, discharges, radio and television transmitters and high frequency circuits. EMI always begins as conductive (current in wires), becomes radiative, and ends as conductive (fields interact with circuitry). Therefore, an electromagnetic shielding is needed for the circuits. A shielded enclosure should be a completely closed conducting surface, i.e. it should have watertight metallic seams and openings. Openings like cooling vents, cable penetration slots, etc. can leak electromagnetic radiation.

10.3.1 Practical Guidelines

Some practical guidelines are:

(a) Twisted pair cable is usually effective upto 1 MHz. It becomes lossy at high frequencies. But it is cheaper and more flexible than coaxial cable. For the transmission of analog inputs over long distances, twisted pair cables with overall or individual pair shield are used.

(b) Coaxial cable has low loss and less variance in characteristic impedance from dc to very high frequencies (>200 MHz). All coaxial cables have shield which must be properly grounded.

(c) Ribbon cable is ubiquitous in instrumentation. The same rules apply to the ribbon cable as those to the twisted pair cable. Figure 10.17 illustrates three possible configurations of ribbon cable. A single return line [Figure 10.17(a)] is good only for low frequency operation. A return plane [Figure 10.17(b)] supports return currents that minimize the inductive loop area and can be used for high frequencies. Paired signal and return lines [Figure10.17(c)] allow high frequencies because they greatly reduce the inductive loop area; twisting each pair is even better.

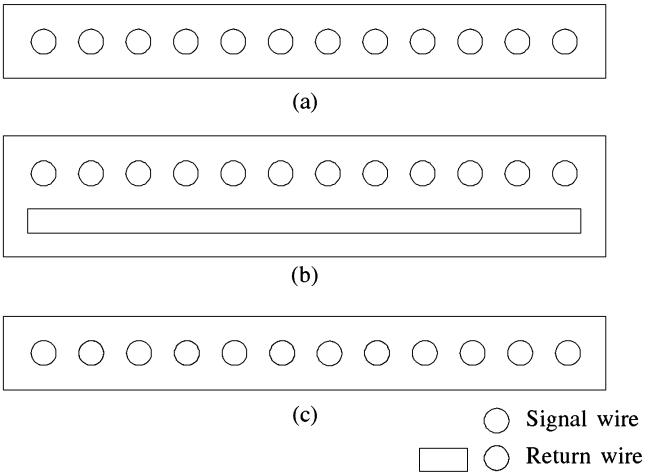


Figure 10.17 Three possible configurations of ribbon cable.

10.3.2 Examples

Ideal Analog Signal Connection

Figure 10.18 depicts ideal conditions for noise-free transmission of a low level analog signal to a distant analog-to-digital converter(ADC), with the signal and shield single-point grounded

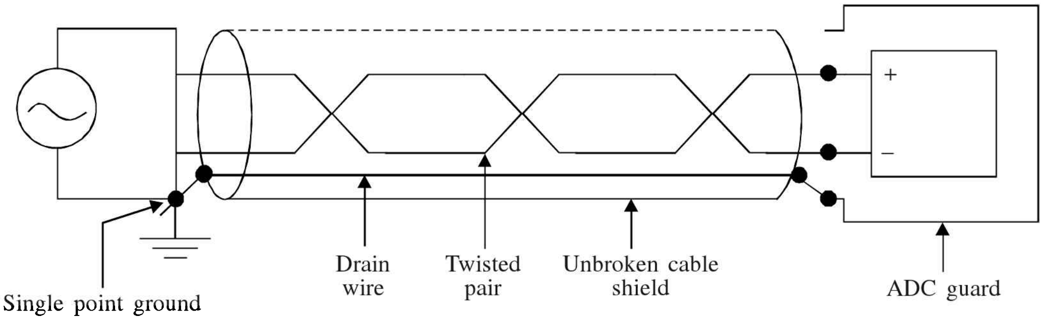


Figure 10.18 Ideal analog signal connection.

once at the source. A twisted pair cable with unbroken cable shield and drain wire (bare copper wire) has been used.

Typical thermocouple analog signal wiring

Figure 10.19 shows a typical wiring of a thermocouple with the termination cabinet. A local junction box has been used for connecting thermocouple wires in the field. The drain wire has been grounded at the local junction box. Other alternatives are field grounding in the thermo-well (within which thermocouple has been placed), near the thermo-well, or in the cabinet.

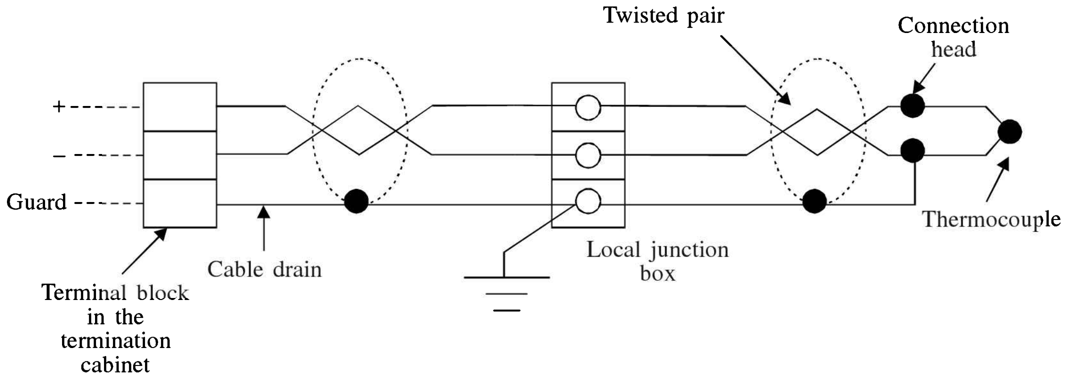


Figure 10.19 Typical thermocouple wiring

Similar wiring arrangement is used for any other type of sensor.

10.4 PROTECTION FROM ELECTROSTATIC DISCHARGE

Electrostatic discharge (ESD) is a discharge at very high voltages (hundreds to thousands of volts) and at very low currents. ESD readily damages sensitive electronics. Instrument containing integrated circuits, specially those using CMOS devices, are susceptible, if not protected.

ESD transfers electric charge in three stages, namely pickup, storage and discharge. Table 10.1 lists some sources of charge. Normally, mechanical rubbing between dry, insulated materials transfers the charge from source to storage. Often the storage medium is a person, who then delivers the damaging discharge. Proximity or physical contact discharges the charge from storage. Several conditions like humidity, speed of the activity and material affect the charge transfer. Typical variation of electrostatic voltages with humidity, are given in Table 10.2.

Table 10.1 Sources of electrostatic discharge

Object or process	*Material or activity*
Work surfaces	• Waxed, painted, or varnished surfaces • Common vinyl or plastic
Floors	• Sealed concrete • Waxed, finished wood • Common vinyl tile or sheeting
Clothes	• Virgin cotton • Wool • Common synthetic personnel garments • Nonconductive shoes • Common clean room smokes
Chairs	• Finished wood • Vinyl • Fiberglass
Packaging and handling	• Common plastic bags, wraps, envelops • Common bubble pack and foam • Common plastic trays, boxes, vials, part bins
Assembly, cleaning, testing and repair	• Spray cleaners • Common plastic solder suckers • Soldering irons with ungrounded tips • Solvent brushes (synthetic bristles) • Cleaning by fluid or drying by evaporation • Temperature chambers • Cryogenic sprays • Heat guns and blowers • Sandblasting • Electrostatic copiers

Table 10.2 Typical electrostatic voltages

Static generation	*Electrostatic voltages (volts)*	
	10% relative humidity	*65% relative humidity*
Walking across carpet	35,000	1500
Walking over vinyl floor	12,000	250
Common plastic bag picked up from bench	20,000	1200
Work chair padded with polyurethane foam	18,000	1500

We can use several schemes, including grounding, shielding, and transient limiters, to protect circuits from ESD. Transient limiters use reverse bias diodes or zener diodes or metal oxide varistors (MOV) for shunting transients away from the circuits. Figure 10.20 shows one such scheme.

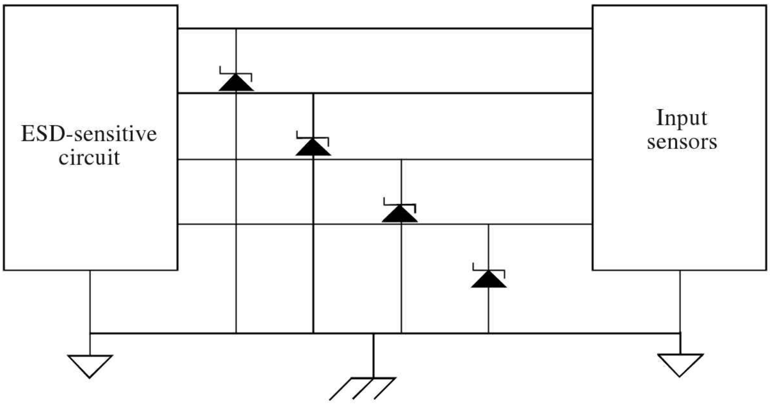
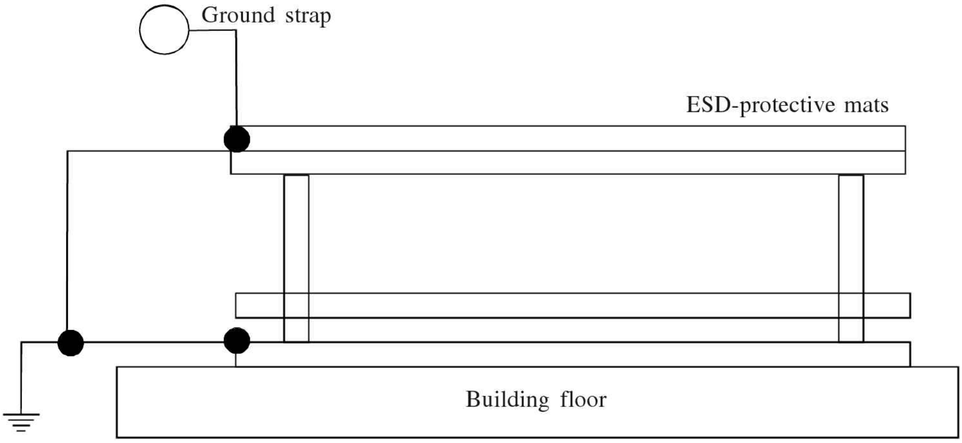


Figure 10.20 Shunting high voltage transients using zener diodes.

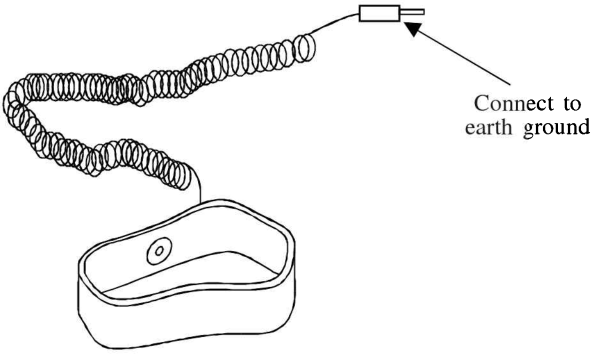
Following precautions make the work area less prone to ESD:

- Using a *static free* workstation and wearing a wrist ground strap (see Figure 10.21).
- Discharging static charge before handling
- Keeping parts in their original containers
- Minimizing handling of components
- Picking up devices by their bodies, not their leads
- Never sliding a semiconductor over any surface
- Using conductive or antistatic containers for storage and transport of components
- Clearing all plastic, vinyl, and styrofoam from the work area.



(a)

Figure 10.21 (*Contd.*)



(b)

Figure 10.21 (a) Typical ESD grounded workbench and (b) ground strap.

QUESTIONS

10.1 What is grounding? Draw the symbols of Common Return, Chassis Return and Ground Return.

10.2 How is a shock hazard created by a leakage path? How would you eliminate a shock hazard from a 1000 volt supply using the appropriate grounding technique? Show diagrammatically.

10.3 What is a ground loop? How is it ascertained in a circuit that a ground loop exists?

10.4 Explain single point grounding and bus bar return.

10.5 What is a ground plane? Find the relationship between the voltage developed in the ground plane and change in the current in the ground plane. How is this voltage reduced?

10.6 Give examples to explain the path of return current in a solid ground plane with and without a slot in it.

10.7 An ADC is to be interfaced between the analog and digital circuits. Illustrate the formation of a ground loop in the interface and the way to remove it.

10.8 Suggest a proper arrangement of grounding of multiple cabinets having separate analog, digital and cabinet grounds.

10.9 How can you ensure noise free transmission of a low level analog signal to a distant I/O card of a PC-based data acquisition system?

10.10 How do you wire a thermocouple with the termination cabinet ensuring all shielding considerations?

10.11 What are the advantages of using twisted wire cables? Explain a typical thermocouple wiring arrangement with the termination cabinet using twisted pair cables.

10.12 What is shielding? Under what conditions shielding is a must?

10.13 Analog signals from the control room of a plant are to be fed to an ADC installed in the field. Suggest an arrangement for noise-free transmission of the signal.

10.14 What is electrostatic discharge (ESD)? What are the sources of ESD?

10.15 How does the electrostatic voltage vary with humidity? Explain with examples.

10.16 What are the methods used to protect circuits from ESD? What precautions should be taken to make the work area less prone to ESD?

10.17 Explain with the help of a diagram an ESD grounded work bench and ground strap.

10.18 How would you save your CMOS devices from high voltage ESD hazard in the laboratory during hand soldering? Show diagrammatically.

Chapter 11

Elements of Design

11.1 INTRODUCTION

Every electronic instrument (or product) is a system of components and interactions and also a part of a large system. Each component is a system unto itself as well. Therefore, the development of electronic instruments is a systematic *system engineering effort*. System components include hardware, software, environmental influences and human operators. System engineering provides a path from concept, through design and test, to delivery and documentation so that the product manufactured is of high quality and has minimal cost.

Systems engineering is an interdisciplinary process which encompasses the integrated activities of evolution, implementation, and verification of all elements that affect the life cycle of a system. Systems engineering encompasses:

- The technical efforts related to the development, manufacture, verification, deployment, operations, support, disposal of and user training for products and processes
- The definition and management of the system configuration
- The development and allocation of functions and requirements to the effective level for systems procurement or construction
- Development of decision criteria for alternative selection
- The translation of the system definition into work breakdown structures
- Gathering and analysis of information for management decision making.

Systems engineering is a process which ensures that:

- The end-product or system is what the customer wanted
- All of the subsystems and parts will fit together
- The completed product will perform, as required, to meet its intended purpose.

We will first study the life cycle of a product and then various aspects of a product with special reference to electronic instruments.

11.2 PRODUCT LIFE CYCLE

The idea of a product life cycle (morphology) acknowledges the fact that designing and selling a product is only part of the story. In fact, every product goes through a series of steps between the time it is first conceived and the time the manufactured product is retired or discarded. Figure 11.1 shows one view of the various phases of a product life cycle.

Need/Opportunity

The product starts with a need or opportunity in the marketplace, for example, digital engine controllers for automobiles, which were virtually required to meet tougher emission controls that could not be handled by mechanical engine controllers. At other times, completely new opportunities exist because of the progress of technology. Introduction of programmable logic controllers in place of electromechanical relays is an example.

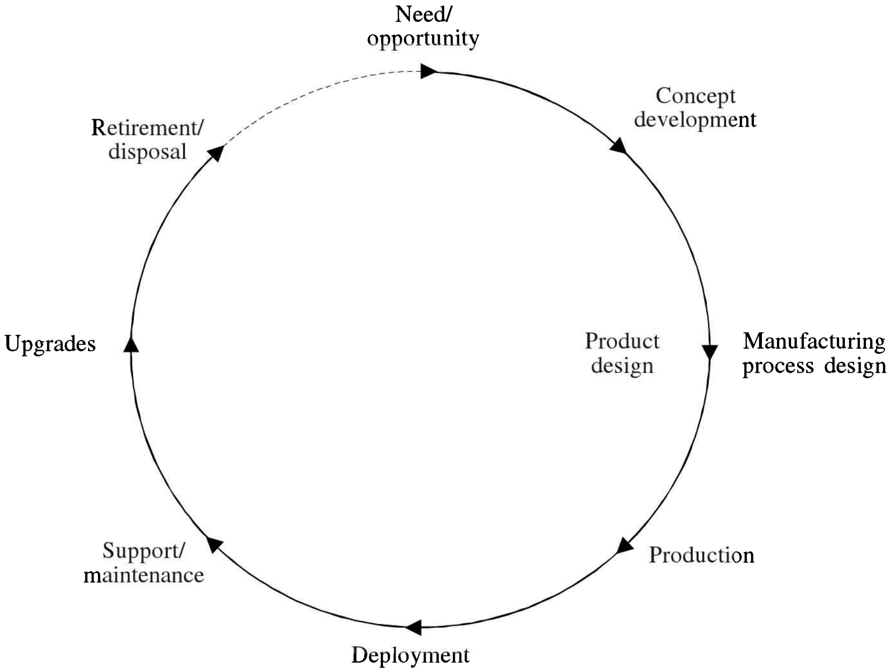


Figure 11.1 The phases of a product life cycle.

Concept development

Once a need or opportunity is defined, a concept for a product is created. The progression of elements within concept development is shown in Figure 11.2.

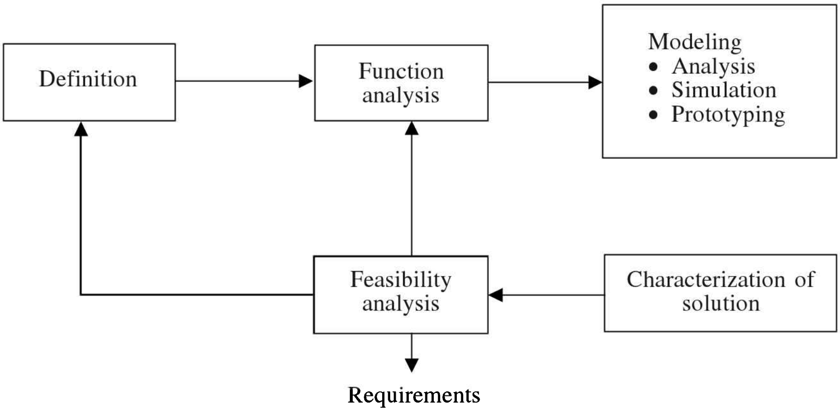


Figure 11.2 Elements and iterative nature of concept development.

The first step is to define the product, taking into consideration the following points:

- Customer objectives
- User needs
- Mission or regions of operation
- Constraints
- Regulations and standards

Functional analysis breaks down operations into more detailed functional components of what must be done. Modeling characterizes possible solutions enabling us to study their feasibility and compare the outcomes. A feasibility study uncovers the limitations and shortcomings of each proposed solution. Sometimes functional analysis and feasibility analysis are interchanged.

Product design

All parts of systems engineering revolve around the synthesis of the solution. Concept development proposes various approaches to design. The requirements structure the framework on which the solution hangs. Integration and testing determine how well the solution fits the requirements.

Within the necessary standards and teamwork approach, development may proceed along one or several lines or methods:

- *Top-down:* The requirements completely drive the design.
- *Bottom-up:* The solution is synthesized from current designs and available technology.
- *Outside-in:* System interfaces drive the design.
- *Inside-out:* The design is driven by developing technology.
- *Hybrid:* A combination of approaches is used.

It is common enough to take the approach of designing something that functions correctly and can be manufactured for a profit. However, with a life cycle approach, it is often better to maximize profit across the entire life cycle, even if that means reaping less profit on the initial equipment sale. For example, a system that is to be upgraded might be designed with higher-priced component sockets or connectors that support multiple installations and removals, rather than less expensive sockets that can only be used a single time.

Considering the deployment requirements of manpower, like training, during the design phase can be quite beneficial. For example, a user interface on a new piece of equipment might be made to mimic the user interface on an older, non-computerized piece of equipment to reduce the need for customer training even though a radically new interface might be slightly more efficient or cheaper to manufacture.

Specific topics identified as important user-oriented concerns that appear in the design process include maintenance, reliability, product safety, convenience in use (human factors), aesthetic appeal, economy of operation, and duration of service. Consumer protection, product safety legislation, and related codes impact designing for service.

Prototyping and testing. Prototypes are created for the designed product, and are bench tested and breadboard evaluated for circuits and devices. These prototypes should be tested at various customer sites for initial evaluation. This is called *beta testing*. The feedback obtained from the sites may pull the design to the intent of the requirements and fit the system to the customer's desires. Testing reduces cost during development. The earlier in life cycle a flaw is detected, the cheaper it is to fix it.

Manufacturing process design

After the product design, the manufacturing processes are designed. In some cases an existing manufacturing process is taken as a fixed parameter, but often it is more efficient to permit changes to the product to accommodate manufacturing needs as well as changes in the manufacturing process to accommodate product needs. If the manufacturing volumes are quite high, it may be worthwhile to modify or even create new manufacturing processes.

Production

Actual manufacturing of the new equipment occurs during the production phase. The design of equipment is frozen at the beginning of production, and design engineering during this phase is limited to correcting problems or developing upgrades. Consideration of production techniques in product design can reap significant profits. For example, automatic component placement machines can only handle a fixed number of different components, so designing to stay within that limit either speeds up production throughput or avoids the cost of buying a second placement machine.

The relationship between design, production and economics is very important for an optimum design. The flow chart of Figure 11.3 illustrates the factors to be considered during these stages.

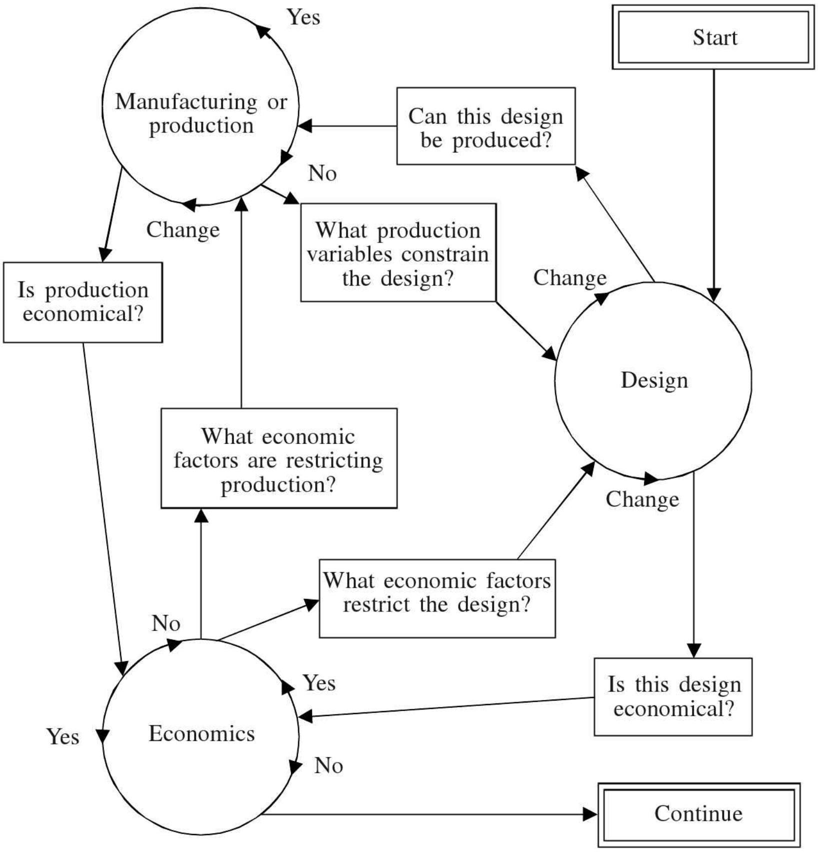

Start
Yes
Manufacturing or production
Can this design be produced?
No
Change
What production variables constrain the design?
Change
Is production economical?
Design
What economic factors are restricting production?
Change
What economic factors restrict the design?
No
Yes
Is this design economical?
Yes
Economics
No
Continue

Figure 11.3 Decision making based upon the interrelationship of design, production, and economics.

Deployment (Fielding)

It includes the activities necessary to initially deliver, transport, receive, process, assemble, install, checkout, train, operate, house, store, or field the system to achieve full operational capability.

Support

It includes the activities necessary to provide operations support, maintenance, logistics, and material management.

Upgrades

The need for upgrades varies dramatically depending on the life expectancy and usage of the product. At the low end, there may be some products in which repair labour is more expensive than the product, or the products may be evolving so quickly that they become obsolete before having a chance to break (for example, digital cameras). On a larger time scale, however, products must be upgraded once or more during their product life. For example, PLCs can remain operational for many years, but are upgraded frequently. Products using software are upgraded very frequently. With equipment that is likely to be upgraded, it is important to design the system in such a way that upgrades can be done cost-effectively, and that system downtime is minimized during upgrade.

Retirement/disposal

Useful life of a product may be determined by actual deterioration and wear to the point at which the product can no longer function, or it may be determined by technological obsolescence. In consumer products, it may come through changes in fashion or taste. Some products may use depleting raw material. In such a case, the product must be designed keeping in mind ease of scrap recovery. Disposal includes the activities necessary to ensure that the disposal of decommissioned, destroyed, or irreparable system components meets all the applicable regulations and directives.

We will now study the factors to be considered during the design process of instrumentation systems.

11.3 CIRCUIT DESIGN

The first step in circuit design is to finalize requirements/specifications of the system. Requirements crystallize the function, operation and performance of the system and give rise to the specifications of the product. Table 11.1 outlines the general types of requirements with specific examples of parameters.

Table 11.1 Outline of requirements

General types of requirements	*Specific examples of parameters*

Performance	• Range
	• Speed
	• Throughput
	• Error rates
	• Size
	• Weight
	• Power consumption
	• Efficiency
	• Test levels
	  • Electromagnetic interference
	  • Vibration and shock
	  • Thermal cycle

Reliability and maintainability	• Mean time between failures (MTBF) • Failure rate • Maintenance downtime

(*contd.*)

Table 11.1 Outline of requirements (*contd.*)

General types of requirements	*Specific examples of parameters*
Human factors and user interface	• Response latency • Number of operations per sequence • Expertise required • Intuitive operation • Ease of use
Safety and failure modes	• Failure mode, effect and critical analysis • Fault-tree analysis • Hazard analysis
Operational regimes and environment	• Duty cycle • Location • Temperature extremes • Stress range
Logistics support	• Maintenance intervals • Personal expertise • Maintenance-task analysis

Example: Specifications of a 4-1/2 digit digital multimeter are given in Table 11.2.

Table 11.2 4-1/2 digit digital multimeter specifications

Measuring functions	ac/dc voltage, ac/dc current, resistance, frequency, diode checking and continuity checks						
Additional functions	Deviation measurement, dBm measurement, dB measurement, comparator, range control, data hold, key lock, last memory and range display functions						
Input characteristics	Accuracy ± % of reading ± digit						
DC voltage	*Range*	*Full scale*	*Resolution*	*Measuring accuracy*			*Input resistance*
	200 mV	199.99 mV	10 mV	±0.03±2			>1000 M
	2 V	1.9999 V	100 mV	±0.05±2			11 MΩ ±5%
	20 V	19.999 V	1 mV	±0.05±2			10 MΩ ±5%
	200 V	199.99 V	10 mV	±0.05±2			10 MΩ ±5%
	1000 V	1000.0 V	100 mV	±0.05±2			10 MΩ ±5%
	Range	*Full scale*	*Resolution*	*Measuring accuracy*			*Input resistance*
				40 Hz to 10 kHZ	10 kHz to 20 kHz	20 kHz to 100 kHz	
AC voltage (true rms)	2 V	1.9999 V	100 mV	±0.4±30	±1±40	±5±100	11 MΩ ±5%
	20 V	19.999 V	1 mV	±0.4±30	±1±40	±5±100	10 MΩ ±5%
	200 V	199.99 V	10 mV	±0.4±30	±1±40	±5±100	10 MΩ ±5%
	750 V	750.0 V	100 mV	±0.1±30 (40 Hz to 1 kHz)			10 MΩ ±5%

(*contd.*)

Table 11.2 4-1/2 digit digital multimeter specifications (*contd.*)

DC current	*Range*	*Full scale*	*Resolution*	*Measuring accuracy*	*Input resistance*	
	200 mA	199.99 mA	10 mA	±0.5±2	About 1.0	
	2000 mA	1999.9 mA	100 mA	±0.7±2	About 1.0	
AC current (true rms)	*Range*	*Full scale*	*Resolution*	*Measuring accuracy*	*Input resistance*	
	200 mA	199.99 mA	10 mA	±1.5±40	About 1.0	
	2000 mA	1999.9 mA	100 mA	±1.7±40	About 0.1	
Resistance	*Range*	*Full scale*	*Resolution*	*Measuring accuracy*	*Open voltage*	
	200 Ω	199.99 Ω	0.01 Ω	±0.1±5	About 3.0 V	
	2 kΩ	1.9999 Ω	0.1 Ω	±0.07±2	About 1.2 V	
	20 kΩ	19.999 kΩ	1 Ω	±0.07±2	About 1.2 V	
	200 kΩ	199.99 kΩ	10 Ω	±0.07±2	About 1.2 V	
	2 MΩ	1.9999 MΩ	100 Ω	±0.1±2	About 1.2 V	
	20 MΩ	19.999 MΩ	1000 Ω	±0.3±5	About 1.2 V	
Frequency	*Range*	*Full scale*	*Resolution*	*Measuring accuracy*		
	5 Hz-1kHz	999.99 Hz	0.01 Hz	±0.02±2		
	10 kHz	9.9999 kHz	0.1 Hz	±0.02±2		
	100 kHz	99.999 kHz	1 Hz	±0.02±2		
	200 kHz	199.99 kHz	10 Hz	±0.02±2		
Diode check	*Range*	*Full scale*	*Measuring accuracy*	*Measuring current*	*Open voltage*	*Display value*
	2 V	1.9999 V	±1.5±2	0.5 to 1 mA	About 3 V	Forward voltage
Continuity check	*Continuity range*		*Reaction speed*		*Continuity check*	
	0 to 100 ± 50 Ω		About 1 ms		Electronic beeper	

ENVIRONMENTAL	
Operating temperature	0°C to 40°C 23°C ± 5°C guaranteed accuracy
Humidity	Less than 80%
POWER REQUIREMENTS	100 to 240 V ac (manual switching) 50/60 Hz
PHYSICAL	
Size (W × H × D)	7-7/8 × 3-3/16 × 10-1/4 in. 200 × 80 × 260 mm
Weight	5.9 lbs (2.7 kg)
SUPPLIED ACCESSORIES	Test Leads Fuse (250 V, 0.5 A, time lag) Fuse (250 V, 3 A, time lag)

The requirements are converted into design. The circuit is designed by considering the general configuration of the system and its specifications. Figure 11.4 shows the general configuration for the circuits within a system.

11.3.1 General Issues

General issues involved in circuit design include the following:

- Selection of technology
- Reliability
- Fault tolerance
- High speed design
- Low power consumption
- Noise and error budgets

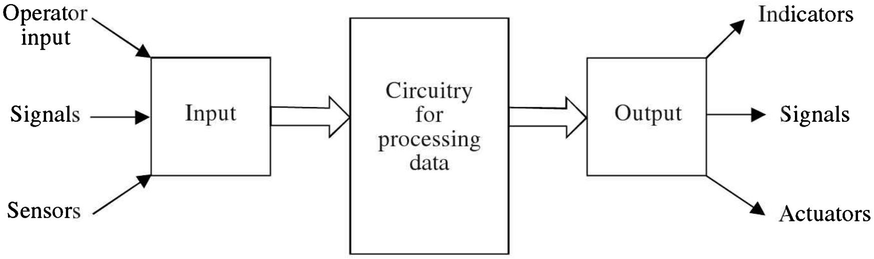


Figure 11.4 General configuration for the instrumentation systems.

Selection of technology

The right choice of the technology normally reduces part counts, PCB size, power, cost and time to market, and increases reliability of a system. Time and effort spent in design increase as the complexity of the function of a system increases. For instance, an ASIC may solve a signal or data processing problem in terms of high throughput and low power, but it will cost more than standard components. If development time merits in our product and cost can be compensated in manufacturing, we can go in for ASIC. Further, we must ensure that proper tools and test equipment are at hand. For example, if our product is based on the microprocessor, we must see that emulators and software tools are available and the designer has good experience in using them.

Reliability

We have to consider factors like how long the product will last, what is the longest duration of continuous operation, what are the environmental conditions, and so on. This will decide reliability of each component and that of the overall system. Reliability can be determined by the use of models and/or prototype tests. Other factors like corrosion, thermal cracks, radiation, vibration, high voltage breakdown, aging, etc. also affect the reliability of the system.

Across a wide variety of mechanical and electronic components and systems, empirical population failure rates have been calculated and a graph (bathtub curve) as shown in Figure 11.5 has been obtained. Note that $h(t)$ in the figure indicates failure rate.

The initial region that begins at time zero when a user first begins to use the product is characterized by a high but rapidly decreasing failure rate. This region is known as the *early failure period* (also referred to as *infant mortality period*). This decreasing failure rate typically lasts several weeks to a few months.

Next, the failure rate levels off and remains roughly constant for (hopefully) the majority of the useful life of the product. This long period of a level failure rate is known as the *intrinsic failure period* (also called the *stable failure period*) and the constant failure rate level is called the *intrinsic failure rate*. Most systems spend most of their lifetimes operating in this flat portion of the bathtub curve

Finally, if units from the population remain in use long enough, the failure rate begins to increase as materials wear out and degradation failures occur at an ever increasing rate. This is the *wearout failure period.*

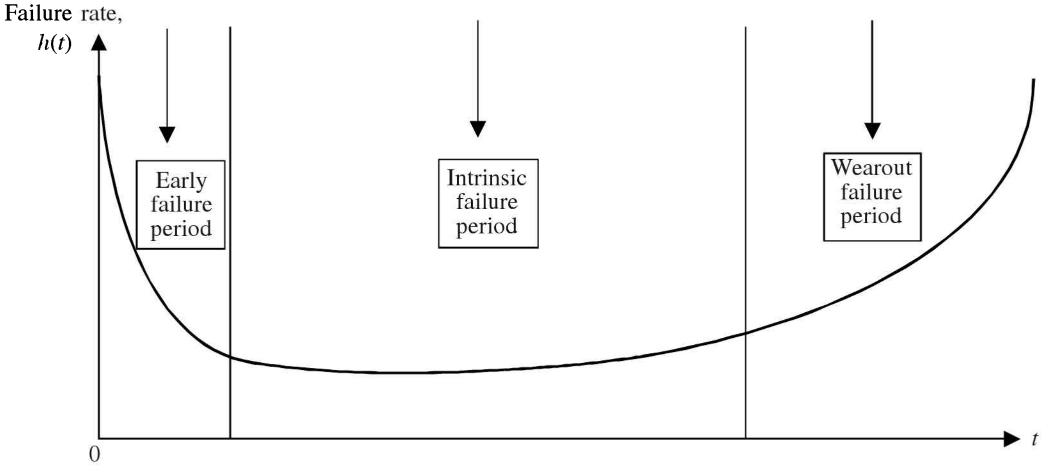


Figure 11.5 The bathtub curve.

Availability

This measure is usually defined as the probability that a system would be working at instant t, regardless of the number of times it may have failed and repaired in interval $(0, t)$. It is denoted by $A(t)$. Availability relates to repairable systems. If the system is not repairable, the definition of $A(t)$ is equivalent to the definition of reliability. Two different measures of availability may be encountered: instantaneous, $A(t)$, and steady state A, as shown in Figure 11.6. The steady-state availability measure is easier to compute but it lacks important information that is captured by the instantaneous availability measure.

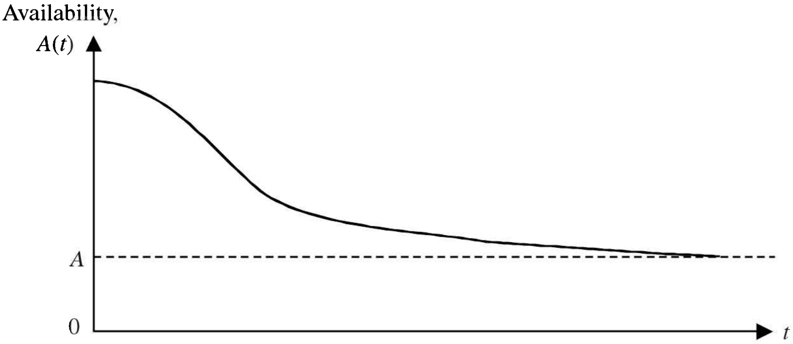


Figure 11.6 Availability curve.

A set of linked measures may be produced for repairable systems. Figure 11.7 shows four of these measures and the links between them.

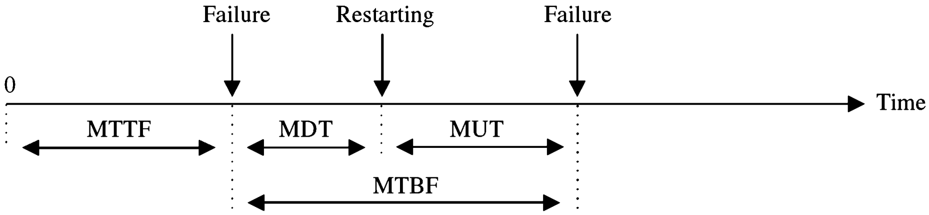


Figure 11.7 Repairable systems measures.

MTTF—Mean Time To (first) Failure
MTBF—Mean Time Between Failures
MUT—Mean Uptime
MDT—Mean Downtime

Another measure sometimes employed is the Mean Time To Repair (MTTR) a component/ system. This is related to the MDT of a system after a failure, but does not include the time to detect the failure and to restart the system after repair.

Steady state availability is given by

$$\lim_{t \to \infty} A(t) = \frac{\text{MTTR}}{\text{MTTR} + \text{MTTF}}$$

Fault tolerance

Fault tolerance is a measure of continuous operation of the system in the event of a problem. It has three distinct areas: careful design, testable architecture and redundant architectures.

Careful design

The probability of failure can be reduced by some design techniques such as:

- Reducing overstress from heat with cooling and lower dissipation design
- Using optoisolation and transformer coupling to stop overvoltage and leakage current
- Using proper grounding and shielding
- Implementing ESD protection
- Proper mounting for absorbing shock and vibration
- Tying down wires and cables that flex frequently and using strain relief
- Preventing incorrect hookup using keyed or polarized connectors

These techniques however may lead to higher cost. So the application of the system should be studied properly and then any or all these factors may be implemented.

Example. A portable instrument using a 9 V battery may not need cooling. But it may not withstand reversed polarity from its battery. We can either use a mechanically polarized socket in the battery compartment or design a rectifier circuit, as shown in Figure 11.8, to provide the proper polarity of the power. But if V_{diode} is significant, other devices can be used.

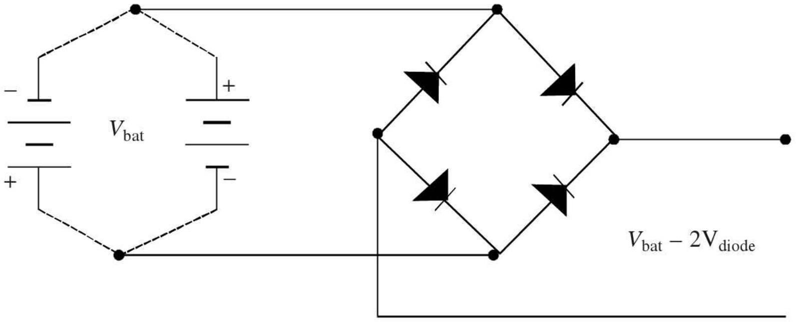


Figure 11.8 Circuit to provide polarity reversal of the battery

Testable architecture. The process of testing and diagnosing failures within a system is another area of fault tolerance. A simple way to do this is by providing test points in the system. Stimulus is applied and the response of components, boards and modules is recorded. Diagnosis is then made manually or using smart devices. A more complex configuration uses dedicated circuitry called built-in test or online diagnosis that tests and diagnose the system without disassembly of the equipment. Such a circuit, however, increases the complexity, and consequently reduces the reliability of the system. User, no doubt, is benefited.

Redundant architectures. These are most complex and use multiple copies of circuitry and software to self-check between functions. These are justified only when downtime for repair and maintenance cannot be tolerated, as for example, in nuclear applications, aviation, etc.

High speed design

Anytime the clock frequency exceeds 1 MHz in a circuit or system, we should consider transmission line techniques, because the harmonics generated by the edges of the clock and signal pulses can be 20 to 30 times the fundamental. When the clock frequency exceeds 30 MHz, wavelength with reference to fundamental is less than 3 m, and the length of some signal lines in the circuit will be a significant portion of the wavelength of the harmonics.

The following two conservative criteria (rule of thumb) may be used to estimate when transmission line effects begin:

(a) Length of the signal path, $l > \lambda/20$, where λ is maximum wavelength of the signal.

(b) Rise time of a signal, $t_r < 4t_p$, where t_p is propagation delay of the signal path.

Careful design can avoid problems associated with transmission lines, including bandwidth limitation, decoupling, ground bounce, crosstalk, impedance match, and timing screw or delay.

Bandwidth limitation. Limiting the bandwidth of the signals within a system is the most effective way to reduce noise, EMI and problems with transmission lines. We can limit the bandwidth either by increasing the rise and fall times of the signal or by reducing the clock frequency (less effective, since rise and fall times contribute higher frequency components). Selecting the appropriate logic family (e.g. $t_r < 4t_p$) sets the edge rates and consequent limit on transmission line concerns. Slower edge rates allow longer interconnections between the circuits.

Decoupling. The switching of digital logic causes transients of current on the voltage supply. These current transients generate voltage transients on the power supply leads through inductive impedance of the circuit. A decoupling capacitor reduces the impedance of the power supply circuit by minimizing the inductive loop area as shown in Figure 11.9. This is particularly true for double-sided PCBs. Multilayer boards with power and ground planes already have very low impedance and do not need as many decoupling capacitors. The decoupling capacitor should occupy the shortest possible path, between the voltage and the ground pins of the ICs, to minimize the inductive loop. We may also use a large capacitor or a ferrite bead at the power entry pins to the circuit board.

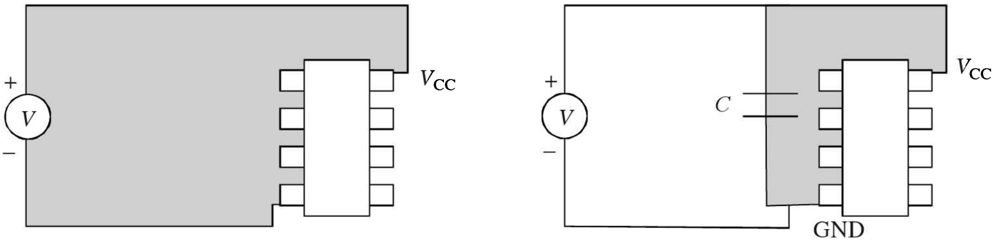


Figure 11.9 (a) Large inductive loop area and (b) decoupling capacitor that reduces the inductive loop.

Ground bounce. Ground bounce is a voltage surge that couples through the ground leads of a chip into nonswitching outputs and injects glitches onto signal lines. The voltage surge is generated when an output (or simultaneous outputs) switches from high to low as shown in Figure 11.10. The switching current from discharging the load capacitance may cause a significant voltage to appear across the loop inductance formed, in part, by the ground lead in the chip. A large enough voltage surge in the ground circuit can cause incorrect operation of referencing other outputs above the input threshold of the driven logic.

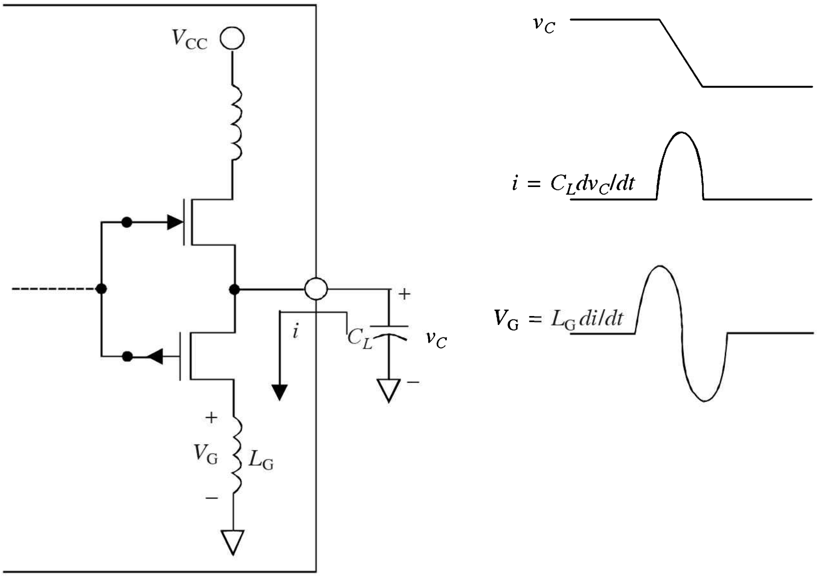


Figure 11.10 Ground bounce.

Asynchronous signals are most susceptible to ground bounce. Further, ground bounce is cumulative. The more the outputs we have switching simultaneously, the greater the magnitude of the ground bounce. Ground bounce can be reduced by reducing the loop inductance (for example, by removing wire wraps and sockets), by reducing the input gate capacitance, and by choosing logic families that either control the signal transition or clause slower fall times.

Crosstalk. Crosstalk is the coupling of electromagnetic energy from an active signal to a passive line. The coupling mechanisms can be capacitive or inductive as shown in Figure 11.11.

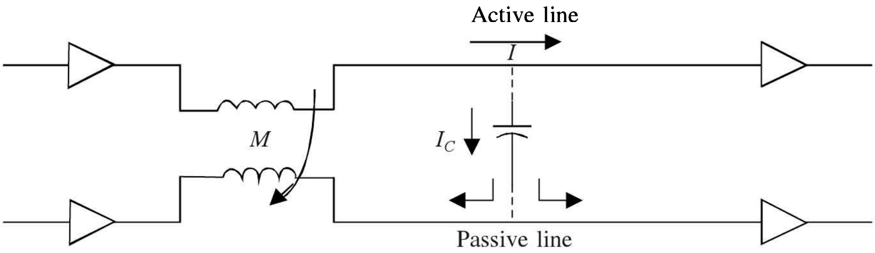


Figure 11.11 Crosstalk.

Crosstalk is a function of line spacing, line length, characteristic impedance, and signal rise time. Decreasing the coupling length and characteristic impedance and increasing the rise time of the signal, all reduce crosstalk. These concerns are often best addressed by good layout.

Impedance matching. Impedance matching makes the source and termination impedances equal to the characteristic impedance of the transmission line. Matching eliminates reflections that cause ringing, undershoot and overshoot in the signal pulses. Impedance discontinuities can be reduced by matching end-point discontinuities, by adding additional series resistances, and reducing stub lengths.

Timing. As clock frequency increases, propagation delays, timing skew and phase jitter can render the design useless. Differences in propagation delay of the clock to different destinations cause the clock to skew, or arrive at different times. Differences in the propagation delay of the rising and falling edges within a device change the duty cycle of the clock, or shrinks it. Crosstalk, ground bounce and propagation delays, all contribute to deviation or phase jitter in the interval between a reference clock edge and the later edges. We must ensure adequate set up and hold time to latch data reliably under these conditions.

Low power design

Low power is used for portability, isolation, battery power, and low heat dissipation. Most of low power consumption instruments use CMOS devices/components where power, P, is given by

$$P = f\,C\,V^2$$

here f is the frequency, C the load capacitance and V the dc supply voltage. Reduction in any of these parameters reduces the power consumption of the system.

Noise and error budgets

Noise generated within each device like Johnson noise ($V_{\text{noise(rms)}} = \sqrt{4kTRB}$), shot noise ($I_{\text{noise(rms)}} = \sqrt{2qI_{\text{dc}}B}$), pink noise [$V_{\text{noise(rms)}} = V_f[0.392 + \log 10(f_{\text{high}}/f_{\text{low}})]$, and interference from power source, radio and TV transmissions etc. must be considered in the circuit design. In addition, errors may be generated due to production, environment, operation of the circuits, etc. Table 11.3 lists the system errors that we should consider and budget for.

Table 11.3 Error budgets considerations for a system

Section	*Component*	*Description of possible errors*
Input	Sensor	Inherent signal noise
	Transmission line	Interference, crosstalk
	Amplifier	Temperature dependency, linearity, CMRR, SNR
	Filter	Variations in component values, band-width, pass-band and stop-band ripple, phase linearity

(*contd.*)

Table 11.3 Error budgets considerations for a system (*contd.*)

Section	*Component*	*Description of possible errors*
Conversion and processing	Sample-and-hold	Aperture jitter, droop
	Analog-to-digital conversion	Quantization error, differential nonlinearity, integral nonlinearity, aliasing
	Digital filtering	Round-off error in calculations
Output	Digital-to-analog conversion	Quantization error, differential nonlinearity, integral nonlinearity, aliasing
	Level shifting amplifier	Temperature dependency, linearity, CMRR, SNR
	Filter	Variations in component values, bandwidth, pass-band and stop-band ripple, phase linearity

Input interface

Input signal to the system may come from another circuit, a transducer or a switch. Most of the signals need some preprocessing, or conversion before the circuit can assimilate them. Switches like push button, toggle, slide, rotary, etc. produce bounce when their position is changed. So proper debouncing circuits should be used. Transducers or sensors may produce voltage or current outputs. Interfacing them with the system may need special pre-processing and isolation circuits.

Output interface

Output signal from the system may drive output devices like LEDs, LCDs, lamps, relays, solenoids, motors etc. LEDs and lamps need current limiting resistances. If large number of lamps change state simultaneously, inrush current from the power supply may increase 10 to 15 times. So power supply design and/or series resistance value should be chosen appropriately. Similarly, changing currents in inductors (coils) in relays and solenoids will induce large voltages across the coils. Series resistances and diodes in parallel to the coils should be used to overcome this problem. Further, system output driving dc motors and stepper motors also need careful design of the driver circuitry.

12.4 CIRCUIT LAYOUT

Printed Circuit Boards (PCBs) are necessary elements in the complex process of developing a functional product. Poorly selected placement and packages of components and routing increase noise, susceptibility, and EMI in electronic circuits. Manufacturing facilities also may change the PCB design. We shall now learn about the procedures and practices adopted for good circuit layout.

Component placement

Thoughtless placement of components complicates the design of the PCB and increases the chances of wrong connections. We should group components according to their characteristics to maintain their correct operation. Following guidelines may be considered for the placement of components:

- Group high current circuits near the connector to isolate stray currents and near the edge of the PCB to remove heat.
- Group high frequency circuits near the connector to reduce path length, crosstalk and noise.
- Group low power and low frequency circuits away from high current and high frequency circuits.
- Group analog circuits separately from digital logic.

Figure 11.12 illustrates these guidelines.

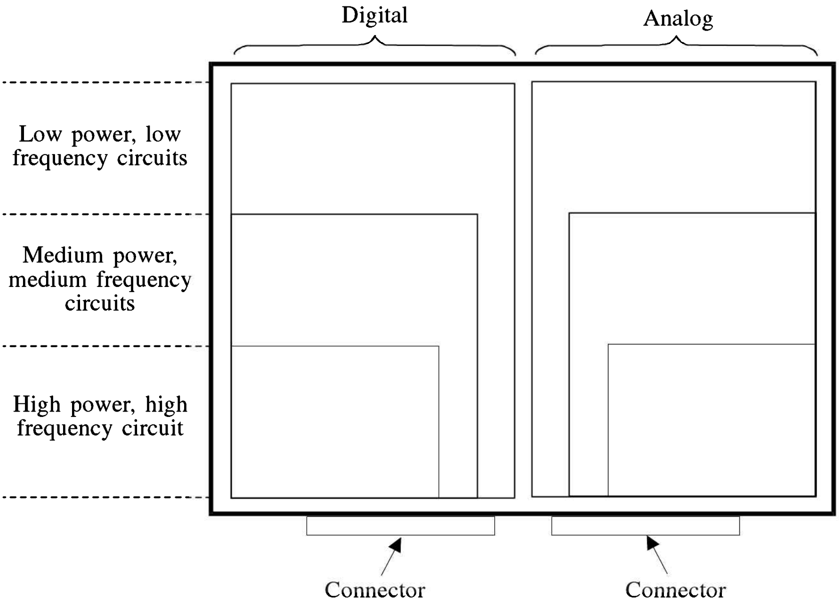

Digital
Analog
Low power, low frequency circuits
Medium power, medium frequency circuits
High power, high frequency circuit
Connector
Connector

Figure 11.12 Grouping of components.

Routing signal traces

Proper routing of the traces (tracks) increases trace density, helps in minimizing the impedance and reduces the effects of noise, crosstalk, and EMI.

Trace density. Smaller boards, allowed by higher trace density, provide flexibility in packaging and reduce the cost of material. But higher densities need precision techniques for production and may degrade signal integrity. Types of devices used also may limit the trace density. For example, number of traces between the pins of DIP (Dual-in-line Package) are more (may be up to 3) compared to packages of SMD (Surface Mounted Devices).

Trace impedance. We should minimize the number of circuits that share the same return path, because voltage drop (caused by current switching) on the return path (ground line) increase system noise. This noise can be reduced by lowering the effective impedance. Choosing the right logic family, and using decoupling capacitors reduce the magnitude of the current pulses. The best way is to have an unbroken return plane. Return path length should be minimum. Also, the lengths of signal and return paths should almost be the same. If the return is longer than the signal, current will have high inductance path and may cause noise spikes in the ground system. Further, proper impedance matching should also be considered.

Reducing crosstalk. Following guidelines are useful to reduce crosstalk while routing signal traces:

- Do not run parallel traces for long distances—particularly asynchronous signals
- Increase separation between conductors
- Shield clock lines with guard strips
- Reduce magnetic coupling by reducing the loop area of the circuits
- Sandwich signal lines between return planes to reduce crosstalk
- Isolate the clock, chip-select, chip-enable, read and write lines because crosstalk occurs in synchronous systems on the pulse edges when data are sampled

These guidelines can run counter to high trace density. We may have to balance noise, EMI and trace density.

Grounds, returns and shields

Grounding and shielding concepts are explained in Chapter 10. Ground provides a reference point for signals. The signal reference should be a single point that is close to the power supply entry to the PCB. Analog and digital circuits should use separate power supply, as for as possible, with the ground points connected to a single point reference.

Keep power and return lines close to reduce loop area and impedance. In multilayer PCBs, separate power and return planes are used. Following guidelines may be used to provide effective shields:

- Use power and return planes with minimum separation.
- Place decoupling capacitors near (or in) IC packages.
- Do not disrupt the power and return planes with slots or traces.
- Route digital traces over digital return (if they must cross into the analog region, keep them short and mostly static, such as gain and multiplexer controls for an ADC).
- Route analog traces over analog return.
- Fill the region between analog traces with copper foil and connect to ground.

We can add shielding by other means too. Mumetal and steel enclosures provide some measure of shielding against magnetic fields. Metal enclosures and cable braiding shield against capacitive coupling of electrostatic fields.

Connectors

Connectors are the mechanical and electrical interface between a PCB and a cable. The metal-to-metal contact within connectors raises several concerns:

- Contact force and resistance
- Gastight fit
- Corrosion
- Life cycle, wear and fatigue
- Keying for polarizing

Corrosion occurs due to high moisture and temperature in the atmosphere. Dissimilar metals also cause galvanic or electrochemical corrosion. Contacts are plated with gold, nickel or tin to improve their resistance to wear, corrosion and oxidation. Gold is an excellent conductor, chemically very stable and best for corrosion resistance. Gold over nickel plating is less expensive and gives surface characteristics of gold while nickel underplating prevents migration of the base metal. Tin is a good conductor and has an excellent solderability. Unfortunately, tin wears off quickly and corrodes easily, and therefore should be used only for contacts that have few insertion/extraction cycles.

While using the connectors, care should be taken to assign the ground pins properly. Interspersing the ground or return pins between signal pins reduces the loop inductance in the signal cables. A ratio of 4 or 6 signal pins per ground pin may be assigned. Figure 11.13 illustrates an example of this assignment with one pin for power (P). G denotes ground pin and S, the signal pin.

Better

S G	S S	G S	S S	S G	S S	G S	S S	S G	S S	G S	S S	S G	S S	G P

Best

G S	S G	G S	S G	G S	S G	G S	S G	G S	S G	G S	S G	G S	S G	G P

Figure 11.13 Ground pin assignment.

Sometimes systems need boards that can be connected while power is on. Online insertion requires careful design of the connectors so that power application to the circuit has an appropriate sequence and prevents latch-up of ICs. If we are using card edge connectors, we can stagger the contact fingers to ensure appropriate sequence of connections when the board is inserted into the system. Figure 11.14 shows this arrangement.

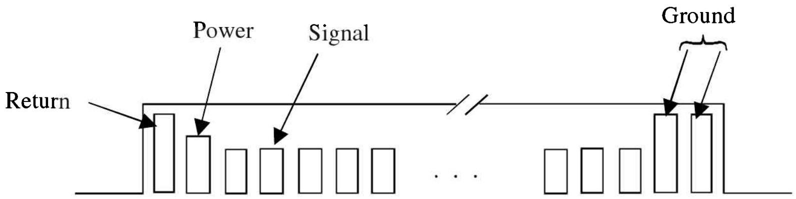


Figure 11.14 Staggering the contact fingers.

11.5 ASSEMBLY AND INSPECTION

Assembly of circuit boards involves placing the components, soldering them and inspecting the solder joints.

11.6 TESTING AND CALIBRATION

Testing can be performed by several ways. Three commonly used methods are the following:

- Scan tests
- Functional tests
- Built-in tests (BITs)

Scan testing stimulates points within a chain of circuitry, records the results and compares them with expected or calibrated data. The circuitry that implements scan testing simplifies testing at the expense of more complexity and real estate on the IC or circuit board.

Functional tests exercise a subset of all the operations that circuit may generate. They require less specialized circuitry but do not check all possible failure modes.

BIT usually dedicates circuitry for testing the remainder of the system. It can implement scan or functional tests. BIT increases the speed of testing at the cost of complexity and lower overall reliability.

Consider the following points during circuit design and layout to accommodate testing at various levels:

- Partition the system, both functionally and physically, so that faults can be detected independently.
- Make circuits testable without the front panel.
- Provide test outputs compatible with the standard test equipment.
- Eliminate warm-up problems or delays to allow rapid testing.

Any measuring and test equipment needs calibration to maintain its quality. Details about the calibration requirements and standards used are given in Chapter 13 on “Standards in Quality Management”.

11.7 POWER DISTRIBUTION

Transmission and distribution of power within an instrument and between modules is an important step in the instrument design. Two configurations used for the power distribution are centralized and distributed power supply. Figure 11.15 shows these configurations.

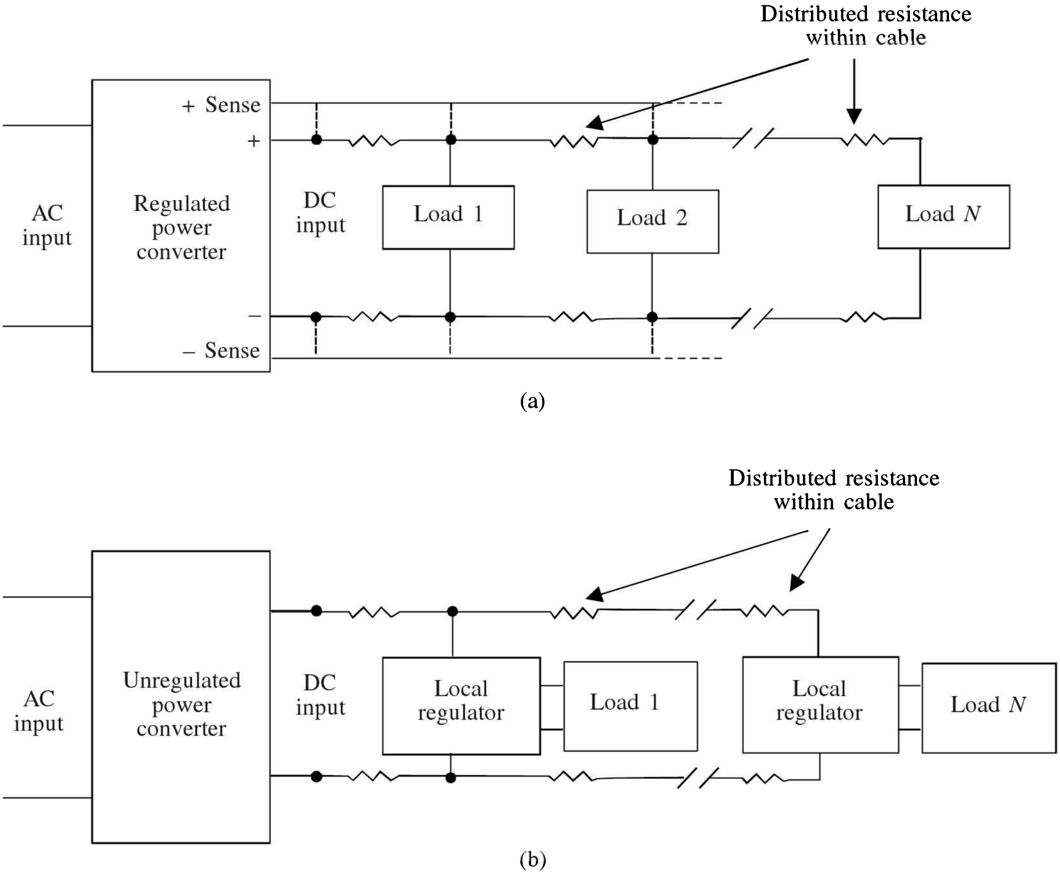

Distributed resistance within cable
+ Sense
+
AC input
Regulated power converter
DC input
Load 1
Load 2
Load N
–
– Sense
(a)
Distributed resistance within cable
AC input
Unregulated power converter
DC input
Local regulator
Load 1
Local regulator
Load N
(b)

Figure 11.15 (a) Centralized power supply scheme and (b) distributed power supply scheme.

The centralized supply converts, regulates, and transmits power from a single location. The distributed power system transmits unregulated power to each subsystem, and a local converter then regulates, the power within each subsystem. Example is a PC based Data Acquisition System where a common power source supplies the power to CPU, I/O cards, hard disk and other devices.

A small instrument with low, well-defined power demands and a fixed architecture with limited expansion favours a centralized supply. A large rack of equipment with multiple modules that are isolated from one another, large expansion capability and wide variations in demand will have a distributed power system. Example is a large distributed data acquisition system where each module has its own dc–dc converter.

For high current requirements, voltage drop in the cable resistance may become appreciable and therefore must be considered while selecting the power supply.

11.8 WIRING AND CABLING

Wires and cables (a bundle of wires) are used to interconnect the modules, connectors, power supply, etc. These are selected based on current carrying capacity, mechanical strength, insulation properties, shielding, and environmental conditions. Following types of cables are used for different applications:

- Coaxial cable
- Fiber optic cable
- Power cable
- Control and signal cable
- Unshielded twisted pair (UTP) cable
- Shielded twisted pair (STP) cable
- Ribbon cable
- Fire Retardant Low Smoke (FRLS) or Low Smoke Zero Halogen (LSZH) cable

Careful routing of cables and tying them down reduces problems arising from vibration, shock and mechanical flexure. Routing and tie-down optimize layout of the wiring for the shortest path, servicing and EMI shielding. Colour coding or wire identification sleeves (ferrules) should be used to aid servicing. General guidelines for cabling are as follows:

- Cabling should be neat, sturdy, and short.
- It should permit easy inspection and testing.
- It should prevent damage to assembled parts and provide sufficient clearance for moving parts.
- It should prevent strain on conductors, connectors, and terminals.
- It should use cable ties 8 to 10 cm. apart.
- It should use grommets (or rounded edges) on holes and penetrations to prevent abrasion.
- It should rout over sharp edges, screws or terminals.
- It should allow reasonable bend radius (10 times the outer diameter of the cable if possible).

Figure 11.16 shows examples of proper cabling and placement of different components, connectors and terminal blocks in the control panels.

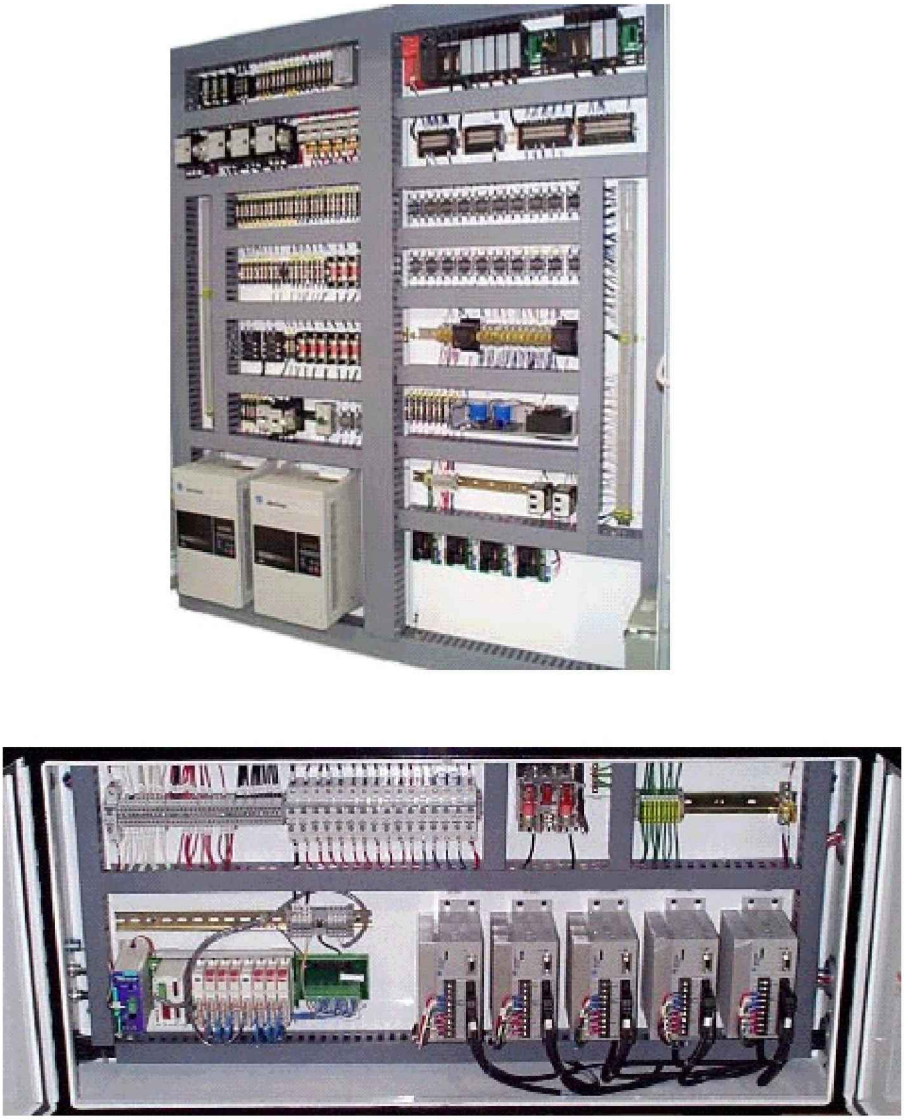


Figure 11.16 Examples of routing of cables and placement of components.

11.9 ENCLOSURES

Factors that affect electronic instruments are the following:

- Temperature and cooling
- Vibration and shock

- Humidity and condensation
- EMI
- Salt spray and corrosion
- Pressure and altitude
- Radiation

Design of an enclosure must take into account these factors. Further, fabricating or selecting an enclosure involves many considerations like:

- Cost
- Size
- Weight
- Material
- Cooling
- Aesthetics
- Ergonomics
- Serviceability
- Regulations and Standards

Physical environment is the biggest factor in deciding an enclosure. See Chapter 15 on "Instrumentation in Hazardous Area" for various standards and coding schemes used for the enclosures.

The materials used affect the weight, cost and performance of the enclosures. Small size plastic enclosures need the trade off between cost, dimensional stability, temperature tolerance, impact resistance, abrasion and chemical resistance. Large cabinets, however, require steel or aluminum sheets and frames for strength and rigidity.

Cooling

Mostly, equipment is designed to operate in one of the following three temperature ranges:

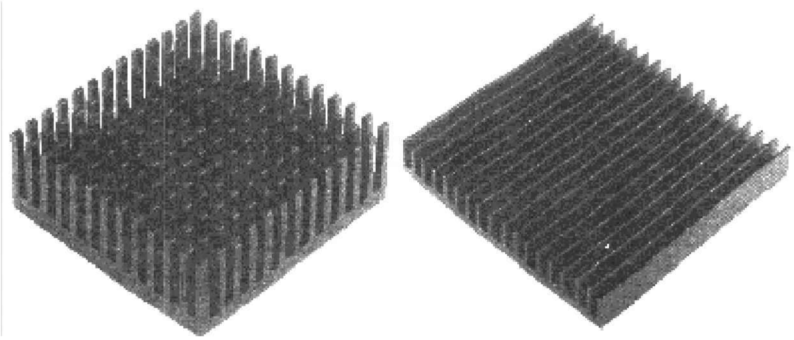
- Commercial: 0°C to 70°C
- Industrial: −40°C to 85°C
- Military: −55°C to 125°C

Chips as they are used in electronic circuits need power in order to operate. The faster the chip, the more clock cycles are operated per second, and more heat is dissipated. Temperature also increases due to the use of power supplies, hard disks, and power transistors in the systems. Further, since industrial systems are operated continuously, even simple PCBs are exposed to high temperature. A proper cooling arrangement is required to control the temperature of the devices and the system as a whole.

There are several ways to achieve cooling:

- Passive cooling
- Forced air cooling
- Passive + forced air cooling
- Active cooling.

Passive cooling. Passive cooling is achieved by means of a heat sink. The dissipated heat spreads over a far larger area and as such leaves the chip itself less hot. Passive cooling helps to cool components that dissipate little heat, since the air surrounding the heat sink is cooler than the heat sink itself, which causes the air to get hotter, and the heat sink to cool down. Figure 11.17 shows two types of heat sinks. Heat sinks with pins have a slight advantage over heat-sinks with fins since heat sinks with pins have another bit of extra surface.



Heat sink with pins Heat sink with fins

Figure 11.17 Heat sinks.

Forced air cooling. Another way to cool equipment is by the means of forced air. A hot chip causes a temperature rise of the air directly around it. With the air getting hotter and hotter around the chip, the chip transfers less and less heat to the already hot air. By forcing air circulation with a fan we can make sure that the air does not get the time to heat up, instead there will always be cold air near the chip. This method is used to cool chips in places where room for a heat sink does not exist. Fans are also used to cool the enclosures and modules. Figure 11.18 gives some examples of forced cooling. Size and speed of the fans should be decided properly to provide adequate air flow.

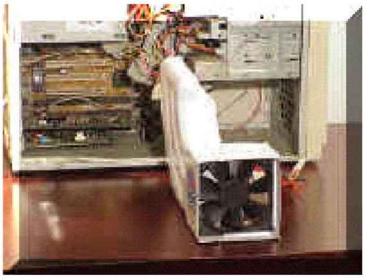


Figure 11.18(a) Fan and a shaft which transports the air through the PCBs.

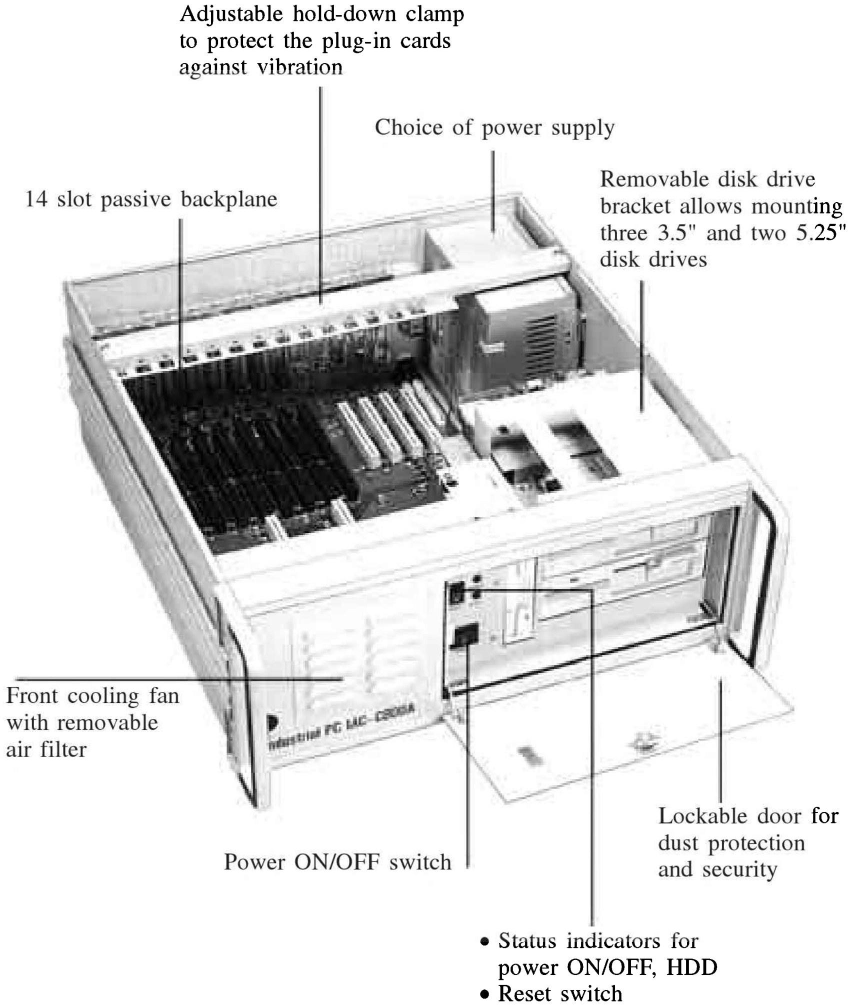
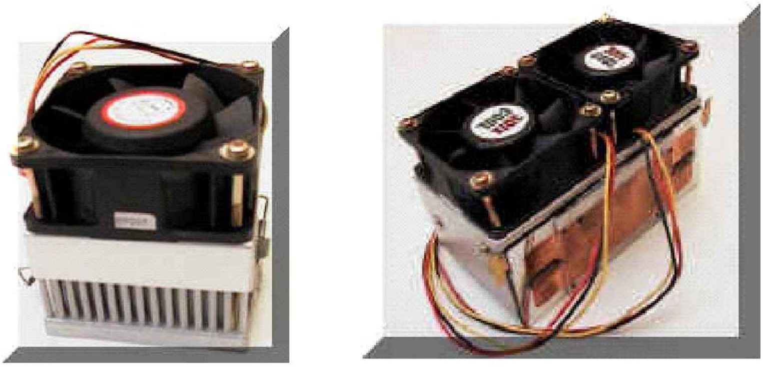

Adjustable hold-down clamp to protect the plug-in cards against vibration
Choice of power supply
14 slot passive backplane
Removable disk drive bracket allows mounting three 3.5" and two 5.25" disk drives
Front cooling fan with removable air filter
Industrial PC IAC-C806A
Power ON/OFF switch
Lockable door for dust protection and security
• Status indicators for power ON/OFF, HDD
• Reset switch

Figure 11.18(b) Industrial PC-based data acquisition system (additional fan with air filter in the front is used in addition to the fan used in the power supply).

Passive + forced air cooling. This third option is most commonly used. It actually involves a heat sink which is used with one or more fans on top of it. The heat sink makes sure that the chip's heat is spread out over a lager surface. Since a heat sink has a large air contact surface, it is much easier to dissipate heat to the surrounding air as a chip-only solution. The function of the fan is like in the other case; it makes sure that the air around the heat sink does not get time to heat up, by continuously supplying cold air. Most of the latest processors come with this scheme. Figure 11.19 shows the arrangement.



(a) (b)

Figure 11.19 Chips mounted with heat sink: (a) one fan and (b) two fans.

Active cooling. Even though the third option is a very good one, there is one problem with it. With forced air cooling we can *never* get the temperature below the ambient temperature. So when we want to get cooler-than-room-temperature cooling, either we use liquid cooling or use a device that actually adds cold instead of removing heat. Such devices are known as thermo electric coolers (TECs), or peltier coolers. They consist of two ceramic plates with positive and negative conductors between them. When current flows through these conductors, one side will get hot while the other chills down. So they actually move the heat from one plate to the other. The temperature difference between the hot and the cold side can be as much as 60°C. Peltier cooling is the only (liquid cooling left aside) way to get a processor to run below ambient temperature at reasonable price.

But there are also a few disadvantages to peltier elements. First of all they draw a lot of current. A peltier can easily consume 60 watts or more. But the amount of power that a peltier unit consumes is not the only problem. Even though one side gets nice and cold, the other side will get very hot. So if we have a peltier cooler, we shall need to add a regular heat sink+fan combo on top of it to cool the hot side of the peltier. This side will get much hotter than the chip itself, so the peltier adds a lot of heat to the system, resulting in a higher in-case temperature.

The design of a liquid cooling system is custom effort and challenging. It increases the cost and reduces the overall system reliability. Failures range from simple leaks to coolant spills and destruction of the circuitry results from overheating.

Aesthetics

The factors which characterize the aesthetics (appearance) of a product are:

- Shape
- Features
- Finish
- Colour and tone

Shape. The shape of the product should conform to functional requirements. It should be simple to manufacture and be in good proportions.

Features. The important features for good appearance include the following:

- Presentation and arrangement of controls
- Presentation and arrangement of displays
- Style and disposition of lettering and symbols
- Choice and arrangement of accessory features such as handles, adjustments and feet.
- Choice and disposition of fixings and fastenings.

Example. Consider the instrument panel display of an analog meter as shown in Figure 11.20(a) and how its appearance is improved in Figure 11.20(b).

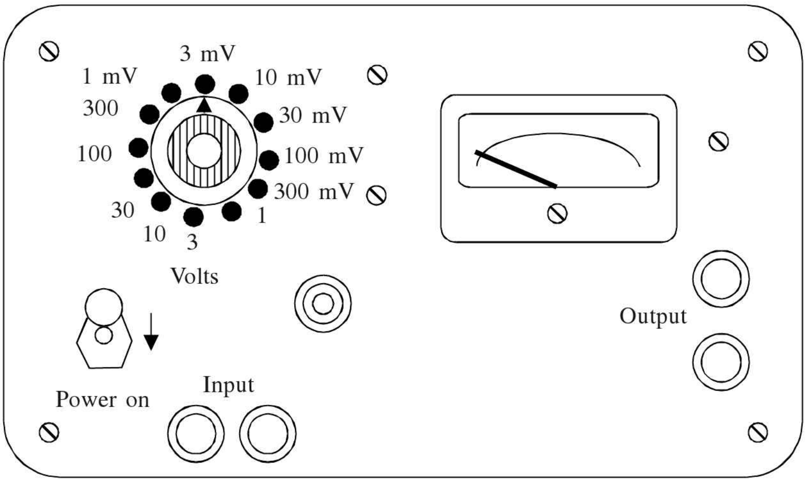


Figure 11.20(a) Instrument panel display with design faults.

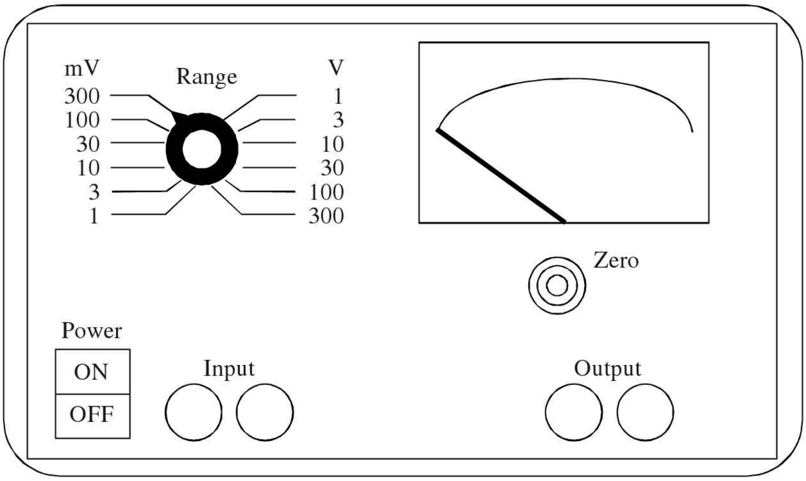


Figure 11.20(b) Instrument panel display with improved design.

Finish. In the choice of surface finish, consideration should be given to the suitability for the application. For good aesthetics, finish should have the following characteristics:

- Protection of base material
- Adequate resistance to wear, abrasion and environment
- No reaction with the substance with which it is likely to come in contact
- Fast colour
- Trouble free and economical to apply
- Correct degree of surface texture or gloss as required by the design
- Easily cleaned.

Colour and tone. Colour (hue) and tone (reflective power) have the effect of adding another dimension to design. These should be chosen to:

- Present a pleasing and non-tiring effect to the eye
- Give contrast where required for the transmission of information
- Give a satisfactory visual balance to the architecture of the product
- Relate the product to its surroundings.

Use of colour is subject to certain restraints. Among these are the association of certain colours with signals (for example, red for danger), the type of use required, such as whether it is continuous or sporadic, restrains brought about by fashion or custom, and requirements either of contrast or harmony with the surroundings. It can be used either to highlight design features or to reduce their effect (for example, products may need to be camouflaged in war time).

Ergonomics

The term ergonomics is derived from the Greek word ergos meaning *work* and nomics meaning *law*. It has two major branches with considerable overlap. One discipline, sometimes referred to as *industrial ergonomics*, or *occupational biomechanics*, concentrates on the physical aspects of work and human capabilities such as force, posture, and repetition. A second branch, sometimes referred to as *human factors*, is oriented to the psychological aspects of work such as mental loading and decision making. The terms ergonomics and human factors are often used interchangeably.

Ergonomics, thus, focuses on accessibility, arrangement, and fit to make a product usable. The benefits of ergonomics are:

- Reduces the risk of producing unfavourable products
- Improves the quality of products
- Improves productivity through reduction of errors, increased accuracy, increased morale and motivation of employees
- Improves safety—lower injury risks

- Low worker turnover and absenteeism rate
- Quick payback time of ergonomic investments

Good equipment design practices and working environment factors provide best ergonomics.

Equipment design practices

- Proper visual displays (e.g. analog and digital indicators, dials, counter wheels, graphics, warning lights, labels, controls, instruction plates, etc.)
- Relevant auditory displays (e.g. speech signals, bell, buzzer, horn, siren etc.)
- Proper size and location of controls (e.g. push buttons, toggle switches, rotary knobs, manipulators, etc.)
- Compatible workplace layout to keep the operator at ease (e.g. operating area dimensions, height of the console, clearance for maintenance, etc.)
- Proper console design and panel layout (e.g. size of the instrument panel depends primarily on the normal arm reach of the human operator, mounting angle of a CRT depends on the position (sitting, standing, or both) of the operator, placement and grouping of controls, etc.)

The working environment

- Good illumination (e.g. to operate the instrument at all times, to repair the system, to avoid glare, to observe numerals and letters, etc.)
- Acoustics and noise control (e.g. to listen to the sound of alarms, unwanted noise and acoustics should be controlled, to design pumps and motors so that noise level is controlled, etc.)
- Temperature and ventilation control (e.g. to control the temperature, humidity and ventilation so that the operator can comfortably operate the instruments and controls, etc.)
- Vibrations control (e.g. reading speeds are reduced by vertical vibrations at low frequency, excess vibrations may be annoying and painful and may affect proper use of the equipment, etc.)
- Safety (e.g. eliminate sharp corners and edges from the enclosures, use danger signs for high voltages, provide proper grounding and shielding, provide locks in the electrical panels, etc.)

Examples

1. Three-position rocker switch for a lamp (Figure 11.21): Its three positions are:

1. Low (I)
2. Off (0)
3. High (II)

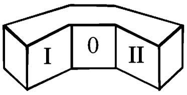


Figure 11.21 Rocker switch.

To switch from Low to High or vice versa, we have to move the switch through the off position. Why it works like this? Would not it not make more sense if its three positions were:

1. Off (0)
2. Low (I)
3. High (II)

That way we would not have to first turn it off to switch from Low to High. It seems like a more natural order. One other problem with this design is that it is difficult to tell when the lamp is set to High vs. Low because it turns completely off between the two settings.

The order of the settings on a control should be natural. We should not have to move a control into the "off" position to move it from one level to the next.

2. View of a good console design (Figure 11.22): Accessibility of the controls on both left and right sides and the required distance between the eyes of the operator and display is taken care of.

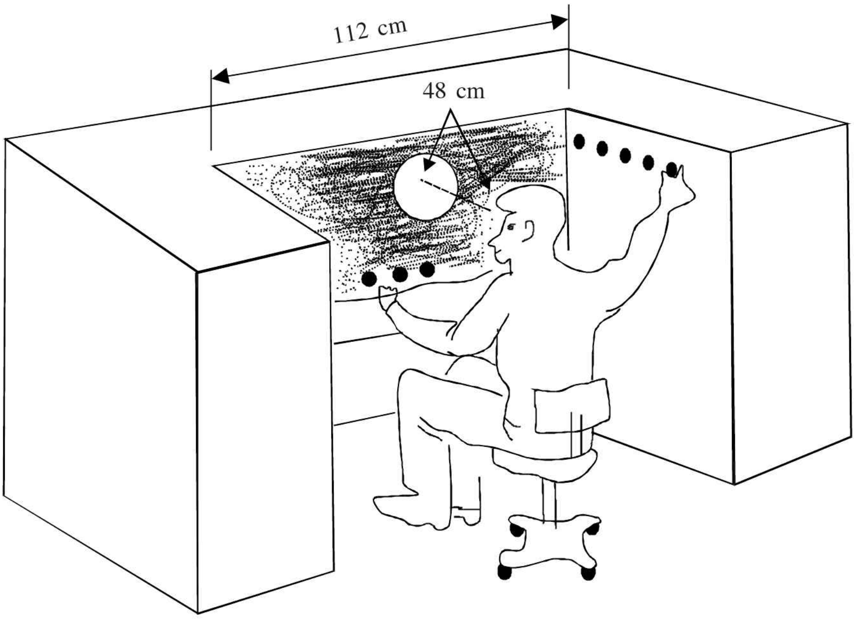


Figure 11.22 View of a console.

Serviceability

The equipment needs operational and post-operational service (maintenance). The serviceability concerns:

- Routine (specified) service operations
- Remedial service in the event of fault
- Overhaul and repair of the damaged or worn equipment/modules.

Following factors must be considered while designing the enclosure:

- All modules/cards are easily accessible from any corner
- Insert/extract levers are used for inserting and removing PCBs
- In case of large panels, proper lighting arrangement is installed
- High voltage areas are isolated and marked
- Standard tools are used for servicing
- Additional power points are available to power any test instrument

Regulations and standards

Different regulations and standards, applicable for various applications, also decide the design of the enclosures. For example, the equipment may be used in following environment:

- Dusty
- Rainy
- Hazardous
- Noisy
- Radiation prone
- Space

For each environment, design criterion will be affected. So, standards and regulations must be consulted before the design is considered.

Example. See Chapter 15 the codes and standards used for enclosures.

11.10 INTEGRATED TESTING

After the component level and fit tests, integration confirms whether the components, modules and subsystems work together as desired. This is the system design verification phase. Figure 11.23 illustrates the general test structure. Some systems may also require additional tests like liquid spills, acoustic noise, humidity, salt spray, etc.

Fit checks. A fit check proves proper clearances between components and mechanisms, cable connections and harness.

End-to-end tests. End-to-end testing is absolutely critical for verifying the operation of any system. We must prepare detailed procedures before integration to avoid mistakes. Step-by-step testing must be performed to ascertain that the complete system operates as designed.

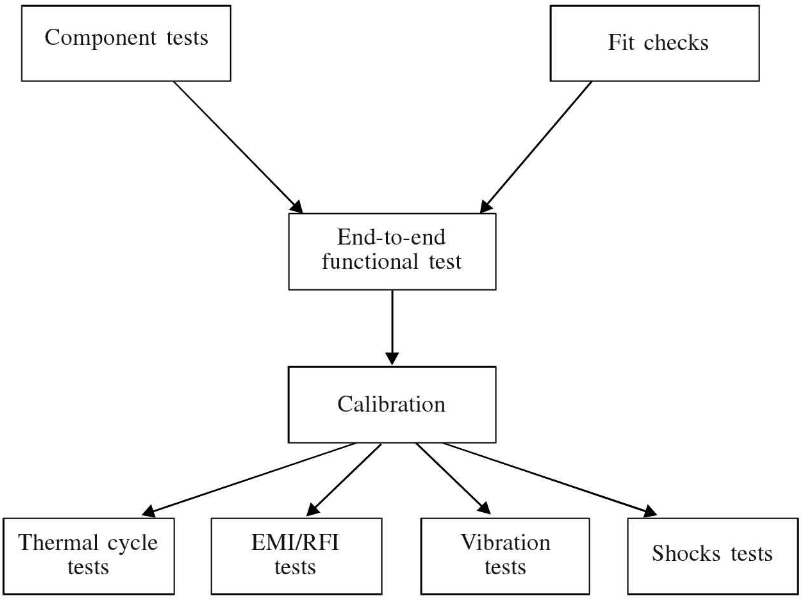


Figure 11.23 General tests structure.

Thermal tests. Thermal tests during integration confirm that the system conforms to environmental requirements. Temperature is cycled under extremes while testing. Some applications may also need concurrent power cycling, vibrations or vacuum.

Vibration and shock tests. If the system is to be used in the industrial environment or in a moving vehicle, then the integrated system should be tested for survival through a series of shock and vibration tests. Each test may deliver a mechanical impulse or vibration in one of the three axes of orientation. Vibration may be either random to present a wide spectrum of frequencies for a duration or swept frequency to identify resonance peaks and determine whether they are below the maximum allowed.

EMI/RFI tests. Independent EMI/RFI tests carried out on the individual modules of the system may not be enough to certify the validity of the tests for the complete integrated system. Subsystems may interact in a complicated manner. So EMI/RFI tests must be carried out on the complete system. Testing laboratories have the capability to perform these tests.

11.11 DOCUMENTATION

Documentation is an integral part to any product. It provides the foundation for understanding the product and promotes its capability and utility. Documentation is necessary for:

- Operation
- Maintenance
- Repair
- Upgrades
- Redesigns
- Fixes
- Legal liability through warnings
- Favourable perception of the product

ISO 9000 guidelines demand records and proper documentation at each stage. See Chapter 13.

Following checklist of questions may help understand the requirements of users to prepare good documentation.

- User capabilities
    - ➢ How familiar are the users with the product?
    - ➢ Do they use similar products?
    - ➢ Are they literate?
    - ➢ Are they trained personnel?
    - ➢ Do they use and understand technical language?
    - ➢ Do they understand charts and diagrams?
- Information needed
    - ➢ Basic introductory material for operation
    - ➢ Educational material for new products
    - ➢ Routine maintenance
    - ➢ Troubleshooting and repair procedures
    - ➢ Explanation of the technology
    - ➢ Are the users in control of the task?
    - ➢ Do they need descriptive detail?
    - ➢ Are they performing upgrades or modifications?
- Cultural concerns
    - ➢ Do the users have a work hierarchy that prevents certain actions?
    - ➢ Do they have any familiarity with technology?
    - ➢ What figures, shapes or colours have favourable connotations?
    - ➢ What figures, shapes or colours are objectionable?
    - ➢ Do the users have preferred writing styles?
- How the manual will be used?
    - ➢ Will it be used to setup and operate the product?
    - ➢ Will it be used only for breakdowns and problems?

- ➢ How often?
- ➢ Do users need to read the entire document or a section?
- ➢ Will the manual lie flat when opened to any page, so that both hands can be used for repairs?
- ➢ Does it need to resist water, grease, mud or food?
- ➢ Will the lighting be good?

Manuals are an integral part of any product because they specify *what*, *why* and *how* to the users. Following are some basic characteristics of the manuals:

- Basic contents
  - ➢ What: prominently note the model number and version
  - ➢ Why: clearly explain the need or theory
  - ➢ How: Detail the operation
- Clear organization
- Simple and to the point writing style
- Modular format
- Clearly illustrated figures and tables
- Detailed schematics
- Table of contents, index and cross references
- Appropriate binding

An introductory section titled "Getting started" often facilitates in understanding the operation of equipment. Many times a user reads only this.

Flow of the manual should be from the general to the specific as given below:

- Table of contents
- Introduction
- General description
- General operation
- Limits of operation
- Maintenance
- Troubleshooting and diagnostics
- Appendices
- Index

As an example, see the manual of an instrument supplied by a good manufacturer.

QUESTIONS

11.1 Describe the various phases of product life cycle.

11.2 Explain the elements of concept development.

11.3 Explain the approaches used for the product design.

11.4 Write five technical specifications of a digital panel meter having 4-1/2 digits display and 100 millivolt full-scale range.

11.5 Draw a flow chart showing the decision-making process based upon the interrelationship of Design, Production, and Economics.

11.6 What are the general issues in the circuit design?

11.7 Explain the bathtub curve.

11.8 Describe the term availability.

11.9 What is meant by fault tolerance? How can the failures be reduced?

11.10 What precautions should be taken for the design to operate at high speed?

11.11 How can the effect of noise, EMI, transients, crosstalk ground bounce and impedance matching be minimized while designing an instrument?

11.12 What are the considerations with reference to grouping of components, routing of signal traces, grounding, shielding, selection of connectors and placement?

11.13 Explain configurations of centralized and distributed power supply schemes used within an instrument and between modules.

11.14 What types of the wires and cables are used to interconnect the modules, connectors, and power supplies? Explain the general guidelines for cabling.

11.15 What are the parameters that affect the performance of an instrument? What are the considerations for the selection/design of an enclosure of an instrument?

11.16 What are the temperature ranges for which most of the instruments are designed? What are the ways to achieve proper cooling in an instrument? How are these types of cooling processes achieved?

11.17 What are the factors which characterize the aesthetics of a product? Explain each factor for a good design.

11.18 Explain the term ergonomics. What are its benefits? Discuss the equipment design practices and working environment factors that provide best ergonomics.

11.19 List a few ergonomical failures noticed by you in the design of a regulated dc power supply, signal generator and CRO used in your laboratory.

11.20 What are the concerns with reference to serviceability of an instrument? What factors should be considered with reference to serviceability while designing an instrument?

11.21 Explain a general test structure for integrated testing of an instrument.

11.22 Why is documentation of an instrument necessary? Prepare a checklist of questions to understand the requirements of users to prepare good documentation.

11.23 What are the basic characteristics of a manual of an instrument? Take manuals of two instruments of different manufacturers and compare their characteristics.

11.24 A 5-1/2 digit digital multimeter continuously gives unstable readings. The signal source is stable. List four environmental and technical factors responsible for instability in a digital display.

Chapter 12

Metrology

12.1 INTRODUCTION

The act of measuring or the process of being measured is called *measurement*. The process involves performing of a predetermined set of operations on a device with the intent of determining some physical property of the device. Metrology is a knowledge field which concerns the measurements. It is the science of measurement. While this definition is stated in its simplest form, it is appropriate because metrology as a science encompasses a vast array of applications and techniques. By simply comparing two or more mathematical quantities, the science of metrology is employed. Metrology can be illustrated by the simple act of reading a household thermostat, and adjusting its temperature accordingly to achieve a desired result. Technically, metrology is the science of measurement associated with the evaluation of its uncertainty, usually expressed numerically as a band within which the actual result is likely to lie.

Metrology is important at all stages of manufacturing—from research and design to the quality of the finished product. This is because:

- Metrology can help guarantee the precision and quality of a widening range of products and production processes.
- Stringent requirements for precision-driven work such as precision component manufacturing need corresponding measurement science and technology which can provide accurate measurement of various parameters and characteristics.
- In the global scale, advanced metrology can enhance competitive capacity of manufacturing industries.

Metrology is particularly important when commercial agreements are based on the ability to make a measurement to mutually agreed standards.

In modern industrial production, goods are mass-produced and further the components and parts of a machine are not made at the same place but at different places, e.g. the components may be made at City 1, City 2, City 3 and assembled say at City 3. At the assembly plant, the different components must 'fit', i.e. no grinding or alteration should be required to make them work together, e.g. the fitting of the piston and cylinder. Indeed now the concept is of replaceable parts. If one component for any reason does not work, an identical component should be available to take its place. The replaced part should not only fit in dimensions but it should also give the same performance. This is possible only if it has the same characteristics including thermal, optical, electrical, magnetic, and other properties. For example, a chip in an equipment, is replaced to get original performance.

The industries cannot be competitive using manual methods, consequently these have been automated. Today the process is controlled by advanced automation, which efficiently stores and processes the information it receives. The information is received in the form of measurement results, and the process is controlled based upon these results. In the industrial countries the high level of know-how and unprejudiced professional attitude towards instrumentation has led people into fast adoption of the latest in technology.

As components age and equipment undergoes changes in temperature or sustains mechanical stress, critical performance gradually degrades. This is called *drift*. When this happens the test results become unreliable and both design and production quality suffer. Whilst drift cannot be eliminated, it can be detected and contained through the process of calibration.

Calibration is simply the comparison of instrument performance (unknown accuracy) to a standard of known accuracy. It may simply involve this determination of deviation from nominal or include correction (adjustment) to minimize the errors. By utilizing metrology, we can estimate the unknown uncertainty with a finite degree of accuracy. Properly calibrated equipment provides confidence that a company's products/services are made according to the required specifications.

Calibration increases production yields, optimizes resources, assures consistency and ensures that measurements (and perhaps products) are compatible with those made elsewhere. And, by making sure that measurements are based on international standards, customer acceptance of one's products is promoted around the world.

Measurements and instrumentation are expanding areas in our automated industry. We need it in order to produce products of the right quality. As the requirements on quality rise, the requirements on accurate measurements rise too. The only possibility to check the accuracy and reliability of the measuring instruments is to calibrate them regularly. The most important weapon in today's competition is the provision of the target quality. Everything is measured by cost, so also is the productivity of calibration. Calibration is an investment with guaranteed returns.

The simple concept behind calibration is that measuring equipment should be tested against a standard of higher accuracy. For any parameter/range we should be able to illustrate this type of hierarchical relationship.

Of course, these calibrations need to be done on a planned, periodic basis with evidence of the comparison results maintained. This record must include identification of the specific standards used (which must be within their assigned calibration interval) and some means of knowing the method used, and other test conditions. By examining these records, it should be possible to demonstrate an unbroken chain of comparisons that ends at the agency responsible for maintaining and developing a country's measurement standards. This demonstrable linkage to national standards, with known accuracy, represents *traceability.*

In fact, it does not stop there because laboratories routinely undertake international comparisons which help to establish worldwide consensus on the accepted value of the fundamental measurement units—without which, there would be little confidence in, for instance, successfully mating a 10 mm screw manufactured in one country with a 10 mm nut produced in another.

Using just a few words we can say that traceability is a declaration stating to which national (and international) standard a certain instrument has been compared. Measurement devices form a kind of hierarchy with international and national standards on the top. There is a network comprising of organizations from the industrialized countries worldwide, in which standards are compared to each other in order to make sure that measurement readings mean exactly the same everywhere.

Traceability could also be expressed as a chain hanging down from the true value. The links represent international standards, national standards, reference standards, working standards (or secondary standards) and process instruments. The true value is thus transferred with a certain accuracy.

We could say that without primary standards, measurement equipment would only show changes and repeat indefinable indications. In case of no traceability to primary standards, the measurements would be just some kind of values. This means that without internationally agreed quantities and without primary standards, everyday technical activity would be impossible. The calibration chain begins with primary standards and ends with process instruments. Thus, when investing in calibration equipment for process instruments it is very important that the supplier guarantees that the calibrators are traceable by showing a calibration certificate indicating against which reference standard and national standard the calibrators have been calibrated.

The foundation of modern metrology is the International System of Units (SI), which is used internationally to define the fundamental units of measurement.

12.2 SI UNITS

The International System of Units (SI) form the basic foundation of modern metrology. These are used to derive other units and are used to define other measurement units, such as may be used locally. The SI units consist of 28 units:

- Seven base units
- Two supplementary units
- Nineteen derivative units

These units are defined in Table 12.1.

Table 12.1 SI units

Base Units

Parameter	*SI unit*	*Abbreviation*	*Definition*
Length	metre	m	The distance travelled by light in a vacuum in 1/299,792,458 s.
Mass	kilogram	kg	The mass of a cylinder of platinum irridium alloy kept in Paris, France.
Time	second	s	The duration of 9,192,631,770 cycles of radiation corresponding to a particular transition of the caesium-133 atom.
Electric current	ampere	A	The current producing a force of 2×10^{-7} newtons per metre of length between two long wires 1 m apart in space.
Thermodynamic temperature	kelvin	K	1/273.16 of the thermodynamic triple point of water.
Luminous intensity	candela	cd	The luminous intensity in a given direction of a source that emits monochromatic radiation at a frequency of 540 × 1012 herz, with a radiant intensity in that direction of 1/683 watts per steradian.
Amount of substance	mole	mol	The amount of substance in a system that contains as many elementary items as there are atoms in 0.012 kg of carbon 12.

Supplementary Units

Parameter	*SI unit*	*Abbreviation*	*Definition*
Plane angle	radian	rad	A plane angle with vertex at the centre of a circle that is subtended by an arc equal in length to the radius.
Solid angle	steradian	sr	The solid angle with vertex at the centre of a sphere that is subtended by an area of the spherical circle equal to that of a square with sides equal in length to the radius.

Derived Units

Parameter	*SI unit*	*Abbreviation*	*Definition*
Frequency	herz	Hz	1/s
Force	newton	N	kg-m/s^2
Pressure	pascal	Pa	N/m^2
Work or energy	joule	J	N-m

(*contd.*)

Table 12.1 SI units (*Contd.*)

Derived Units

Parameter	*SI unit*	*Abbreviation*	*Definition*
Power	watt	W	J/s
Electric potential	volt	V	W/A
Electric resistance	ohm	ohm	V/A
Quantity of electricity	coulomb	Q	A-s
Electric capacitance	farad	F	C/V
Conductance	siemens	S	A/V
Magnetic flux	weber	Wb	V-s
Magnetic flux density	tesla	T	Wb/m^2
Inductance	henry	H	Wb/A
Celsius temperature	degree	°C	K
Luminous flux	lumen	lm	cd-sr
Illuminance	lux	lx	lm/m^2
Activity	becquerel	Bq	1/s
Absorbed dose	gray	gy	J/kg
Dose equivalent	sievert	Sy	m^2/s^2

12.3 TESTING, COMPATIBILITY, CALIBRATION, AND TRACEABILITY

12.3.1 Testing

The process of measurements performed on the components by the manufacturers or by any other agency is called *testing*. This is an important part for maintaining quality of the products. The items produced are tested for their performance and quality. Normally, testing is performed at the location of the manufacturer.

Tests should be realistic, reproducible, and give results capable of ready interpretation. The equipment used and/or the choice of test method determines the extent of the tests, and test conditions specify the ensured reproducibility.

Testing is used in four applications:

- Product evaluation (especially new products)
- Routine monitoring of production quality
- Investigation of faults or complaints
- Research

12.3.2 Compatibility

If parts of any product are manufactured at many locations, the test instruments at each place should give the same result within specified limits. This ensures compatibility of measurements among themselves.

12.3.3 Calibration

The process of inspecting and maintaining measuring equipment is referred to as a calibration program. Calibration includes the comparison of measuring equipment to a standard, and possible adjustments based on these activities. The calibration program should ensure that measurements made in the design, manufacture, and testing of a product meet the necessary accuracy and reliability requirements. Calibration of measuring and test equipment may be referred to as measurement assurance. It is important that product quality tests are performed with measuring and test equipment that is capable of making accurate and reliable measurements.

The overall purpose of a measurement assurance program is to assist in ensuring that product quality requirements are met. The goal of the manufacturer is to meet or exceed the quality requirements in the most cost-efficient manner possible. Measurement testing is a cost-intensive process, and it may take time before the benefits outweigh the costs. It would not be cost-efficient to calibrate each measurement device everyday, nor would it be a good practice to calibrate or test each device only once per decade. Equipment that has not been calibrated regularly may fall out of specification, and product quality may suffer. The key is finding a balance between the lowest operating cost and the highest quality product.

ISO 9001 and ISO 9002 requirements for calibration

The ISO 9001 and ISO 9002 quality assurance systems require producers to state what measurements are used during the process, what level of error tolerance is allowed for these measurements and which measuring instruments are used for the measurements.

The adaptation of a Quality Assurance System is the most economic way to ensure continuous optimal quality production corresponding to the as new quality. The same applies to the measurements. Every measurement has its own optimal accuracy, which should be already predetermined when measuring instruments are purchased. When specifying the optimal accuracy, the measuring instruments can be divided into three categories of importance, for example A, B, and C each or which have their own error tolerance limits.

The B class is a so-called general class, where most of the plant's instruments belong. The C class contains the so-called secondary measurements, which do not have a big influence on the production quality. The A class is a special quality class containing the most important and demanding measurements. The classification gives the measurements an accuracy target. The most accurate measurings should naturally be examined closer and their tolerance limits should also be specified very carefully.

The quality assurance systems require the producer to calibrate the measuring instruments regularly using adequate calibration instruments which are traceable to nationally recognized standards. The producer must rely on a calibration procedure to re-ensure that the calibration of the measuring instruments does not run out of date. Further, the instruments' calibrations must be documented.

It has been demonstrated that the measuring instruments' accuracy weakens during use. The only way to keep track of the measurements' actual accuracy is to verify, calibrate and adjust the measuring accuracy to the original level.

The calibration programme

The basic elements of a calibration program are:

- Testing and calibration procedures are documented.
- Tests to check calibration are performed on a scheduled basis. The schedule should be readily accessible.
- Adjustments to the equipment are made based on test results.
- Confirmation labeling should indicate when the equipment is next due for calibration.
- Measurement equipment that is deemed not to require calibration should be identified as such.
- The tests performed should verify that the measuring equipment used can provide the necessary accuracy and repeatability.
- A system to provide evidence of traceability to national or international standards is used. Certificates of traceability should be available for each standard.
- Calibration records are maintained.

Depending on the results of previous calibrations, the calibration interval may be shortened to ensure continued accuracy or may be lengthened when warranted. The purpose of having a calibration program is to maintain, and in some cases, improve product quality and meet customer requirements.

Typical areas and applications, which require that instruments be calibrated, include:

- Product quality
- Designed experiments
- Environmental testing
- Testing against predetermined specifications
- Safety
- Failure mode analysis

12.3.4 Traceability

Traceability is a term used to describe the unbroken chain of comparisons that qualify measurement equipment to national or international standards. A quantity is measured with measurement equipment. Measurement equipment is calibrated with standards. A standard can be calibrated with a primary standard. The primary standards which are sophisticated big and bulky, are difficult to install since they require an extended experimentation to realize them. In everyday life we require convenient and reproducible standards, which can be transported so that it will help us to achieve accuracy as close as possible to the primary standards. These are called "Transfer Standards". National laboratories help in realizing them so that these can be given to the laboratories lower down in hierarchy. These transfer standards assure the accuracies of their laboratories and form the link in traceability to the national standards. Traceability certificates should state which standards have been used to calibrate the instruments. Primary standards are standards of the highest accuracy available. Figure 12.1 illustrates the hierarchy of the traceability chain. The National Institute of Standards and Technology (NIST) is the standards organization in the United States. Its equivalent in India is National Physical Laboratory (NPL).

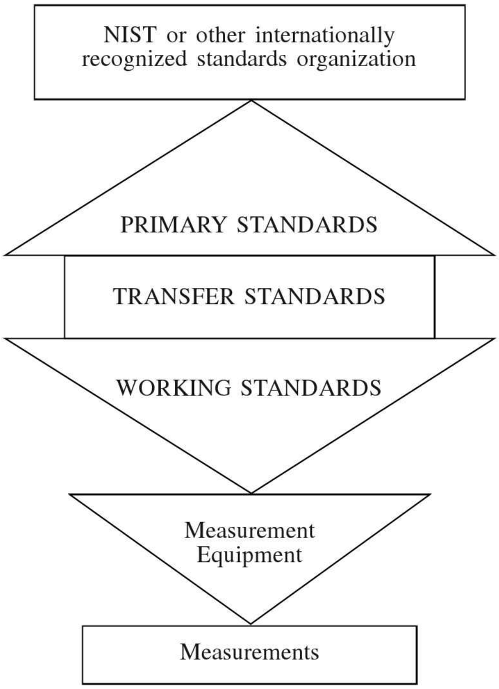


Figure 12.1 Hierarchy of traceability chain.

Figure 12.2 shows the various echelons for the traceability of accuracies in electronics measurement under the National Calibration Service (NCS).

It is clear from Figure 12.2 that the accredited organizations at Echelon 2 and 3 should keep on their own, reliable instruments of specified accuracies. They should get their instruments calibrated periodically by a superior laboratory at Echelon 1 which has the capability of one order of magnitude higher accuracy. This may be the apex laboratory for the

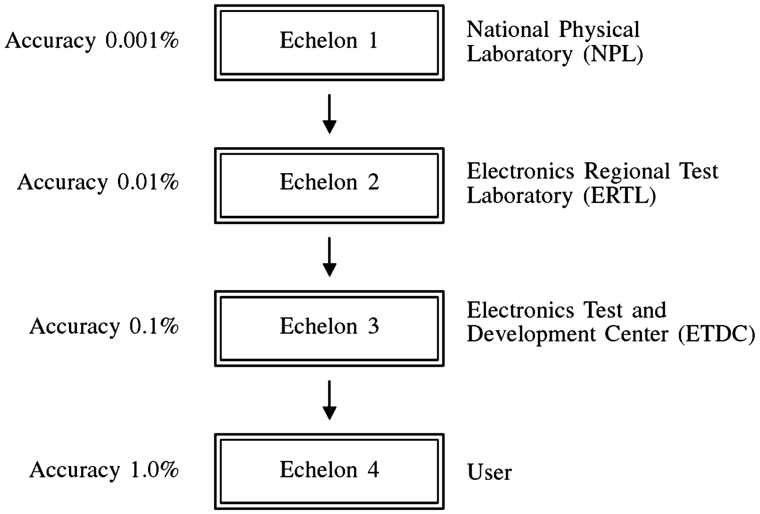

Accuracy 0.001%
Echelon 1
National Physical Laboratory (NPL)
Accuracy 0.01%
Echelon 2
Electronics Regional Test Laboratory (ERTL)
Accuracy 0.1%
Echelon 3
Electronics Test and Development Center (ETDC)
Accuracy 1.0%
Echelon 4
User

Figure 12.2 Echelons for the traceability of accuracies in measurement.

country, e.g. NPL (or BARC for ionizing radiations) for India. The apex laboratory creates the accuracy by suitable experimentation and observation. This requires high scientific competence.

Thus the accuracy of different organizations and laboratories in India should be in stages of hierarchy traceable to NPL.

For international trade and commerce, it is not enough that a particular country has a traceable system. In order that the measurements made in one country should agree (within specified limits) with the measurements made in another country, it is essential that the accuracy of each country should be traceable to the international accuracy by a suitable procedure.

By international agreement a treaty called "Convention du Metre" was signed in 1875 and the signatory countries have accepted the use of SI units of measurement. This system now universally accepted. By the agreement, a permanent international scientific body of weights and measures "Bureau International des Poids and Measures" abbreviated BIPM has been established at Paris. The signatory countries have a body called CGPM (General Conference on Weights and Measures) for the overall running of BIPM. The accuracy of all national laboratories should be traceable to the international accuracy maintained at BIPM in Paris. The international laboratories and BIPM work in cooperation to maintain the international accuracy.

12.3.5 Measuring Equipment Cycle

Figure 12.3 shows the measuring equipment cycle involved in the testing and calibration of any manufactured/purchased equipment.

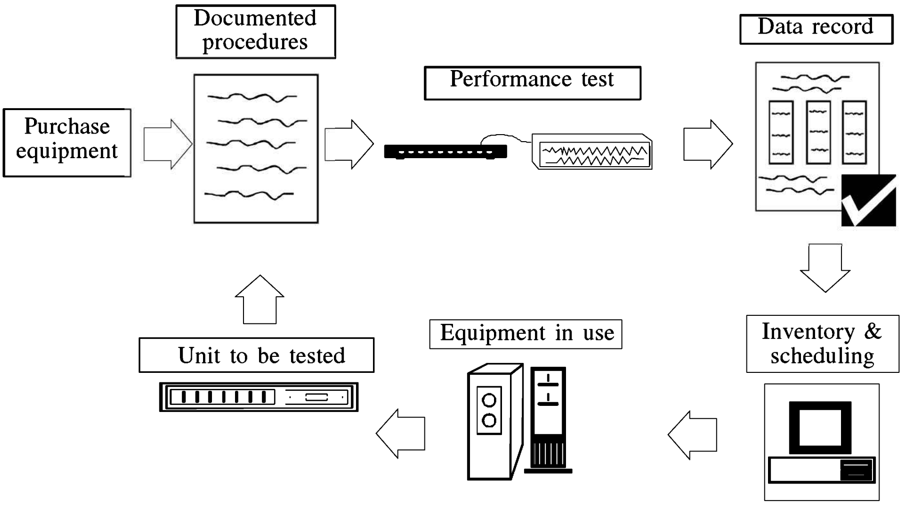


Figure 12.3 Measuring equipment cycle.

12.3.6 Measurement Error and Measurement Uncertainty

The measurement error is the estimated amount that a measured value differs from its true value. Measurement errors can stem from equipment, operators, test design and various other factors. Many are difficult to identify and quantify. To increase the accuracy of measurements, errors must be minimized. In order to develop uncertainty statements, suspected measurement errors are assigned estimated probability values.

It can never be certain that the measured value of a reading is the true value. Measurement readings are estimates of true values. Measurement uncertainty may be defined as the probability that a reading will fall in the interval that contains the true value.

Thus uncertainty is a figure of merit associated with the actual measured value; the boundary limits within which the *true* value lies. Contributors to this “potential for inaccuracy” include the performance of the equipment used to make the measurement, the test process or technique itself, and environmental effects. Additional imprecision may result from behaviour of the phenomenon or item being measured. A skilled metrologist will assess and combine these various components in an uncertainty budget. To prove that a product complies with specification (or does not), the uncertainty must be less than the unknown's specification.

Uncertainty statements consist of two parts, an error value and a corresponding level of confidence. An uncertainty statement such as ±0.1 units, 2 sigma confidence level means that the author of the statement is 95% sure that the true value is not different than the measured reading by more than ±0.1 units. The error value determines the range or interval that contains the true value. The error in this example is ±0.1 units. If the nominal value is 2.0 units, the confidence interval is 1.9 to 2.1 units. This interval may also be referred to as the 95% confidence interval. Uncertainty values are also time dependent. If the confidence level is held constant, the magnitude of the error tends to increase with time. Uncertainty statements normally apply to the end period of the manufacturers recommended calibration interval.

Classification of components of uncertainty

In general, the result of a measurement is only an approximation or estimate of the value of the specific quantity subject to measurement, that is, the measurand, and thus the result is complete only when accompanied by a quantitative statement of its uncertainty.

The uncertainty of the result of a measurement generally consists of several components which may be grouped into two categories according to the method used to estimate their numerical values:

A. Those which are evaluated by statistical methods

B. Those which are evaluated by other means.

There is not always a simple correspondence between the classification of uncertainty components into categories A and B and the commonly used classification of uncertainty components as *random* and *systematic*. The nature of an uncertainty component is conditioned by the use made of the corresponding quantity, that is, on how that quantity appears in the mathematical model that describes the measurement process. When the corresponding quantity is used in a different way, a random component may become a systematic component and vice versa. Thus the terms *random uncertainty* and *systematic uncertainty* can be misleading when generally applied. An alternative nomenclature that might be used is

"component of uncertainty arising from a random effect,"

"component of uncertainty arising from a systematic effect,"

where a random effect is one that gives rise to a possible random error in the current measurement process and a systematic effect is one that gives rise to a possible systematic error in the current measurement process. In principle, an uncertainty component arising from a systematic effect may in some cases be evaluated by method A, while in other cases by method B as may be an uncertainty component arising from a random effect.

The difference between error and uncertainty should always be borne in mind. For example, the result of a measurement after correction can unknowably be very close to the unknown value of the measurand, and thus have negligible error, even though it may have a large uncertainty.

The evaluation of uncertainty by the statistical analysis of a series of observations is termed a Type A evaluation (of uncertainty). The evaluation of uncertainty by means other than the statistical analysis of a series of observations is termed a Type B evaluation (of uncertainty).

Type A evaluation of standard uncertainty may be based on any valid statistical method for treating data. Examples are, calculating the standard deviation of the mean of a series of independent observations; using the method of least squares to fit a curve to data in order to estimate the parameters of the curve and their standard deviations, and carrying out an analysis of variance (ANOVA) in order to identify and quantify random effects in certain kinds of measurements.

Type B evaluation of standard uncertainty is usually based on scientific judgment using all the relevant information available, which may include:

- previous measurement data

- experience with, or general knowledge of, the behaviour and property of relevant materials and instruments
- manufacturer's specifications
- data provided in calibration and other reports
- uncertainties assigned to reference data taken from handbooks.

Because the reliability of evaluations of components of uncertainty depends on the quality of the information available, it is recommended that all parameters upon which the measurand depends be varied to the fullest extent practicable so that the evaluations are based as much as possible on observed data. Whenever feasible, the use of empirical models of the measurement process founded on long-term quantitative data, and the use of check standards and control charts that can indicate if a measurement process is under statistical control, should be part of the effort to obtain reliable evaluations of components of uncertainty. Type A evaluations of uncertainty based on limited data are not necessarily more reliable than soundly based Type B evaluations.

12.3.7 Determining Instrument Capability

By quantifying and combining possible sources of error throughout the chain of traceability, it is possible to estimate and compare instrument capability with test requirements. The uncertainty of a reading should be small enough to have a negligible effect on the result of the measurement. A common method is to calculate the test uncertainty ratio (TUR), sometimes referred to as the test accuracy ratio (TAR). The uncertainty of the standard(s) used for calibration is combined with instrument uncertainty in order to derive a total measurement uncertainty statement for instrument output. When an instrument is used to make a measurement, the total uncertainty, at a specified confidence level, is compared to the allowable tolerance of the measured value in order to obtain a TUR. NIST recommends a confidence level of ±2 sigma or 95%. TURs are used as a measure of instrument capability. Current industry practices usually accept ratios of 4 to1 or greater, however, 10 to 1 or greater is desirable.

$$\text{TUR} = \frac{\text{tolerance of measured value (TMV)}}{\text{uncertainty of test standard (UTS)}}$$

where the confidence level of the uncertainty of the test standard (UTS) is 95%, the tolerance of the measured (TMV) value is the allowable test limit of the quantity being tested, the uncertainty of the test standard includes the uncertainty of all standards used in the traceability chain.

Example 12.1

A company has set internal goals for achieving at least 4 : 1 test uncertainty ratio (TUR) for all calibration tests.

The tolerance of a caliper is ±0.1 mm (TMV = 0.1). The uncertainty of the gauge blocks used to test the caliper is ±0.02 mm at a 95% confidence level (UTS = 0.02). Are the gauge blocks acceptable for testing calipers?

Solution

$$\text{TUR} = \frac{\text{TMV}}{\text{UTS}} = \frac{0.1}{0.02} = 5$$

The TUR of 5 represents a 5 : 1 ratio. This ratio is above the acceptable limit set by the company. The gauge blocks chosen for this test have been deemed as acceptable for testing calipers.

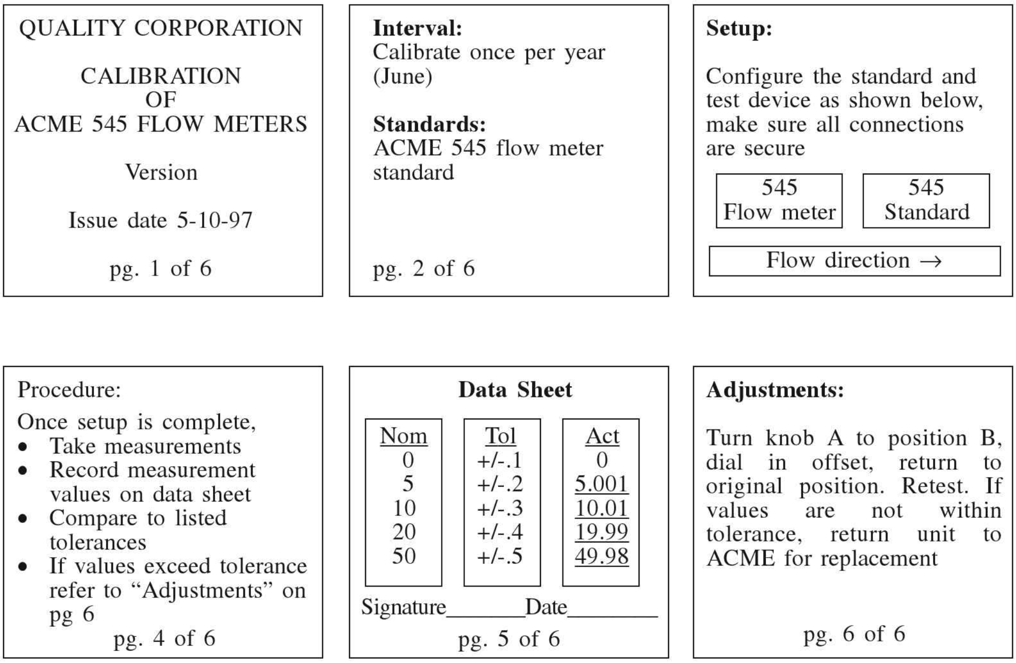
12.3.8 Meeting Test Specifications

Measurement acceptability is determined by comparing values of the unit under test (UUT) to standard values. A nominal value from the UUT is compared to an actual value determined from the standard. The measurement passes if the difference between the two values is less than that specified by the tolerance. If the difference is greater than the tolerance, then the measurement fails and some type of corrective action occurs such as adjustment, repair, or replacement of the UUT. After corrections have been made, testing is again required to determine if the corrections were effective. When testing a measurement device, several values throughout the range of operability or use are usually chosen to be tested. Single checkpoints can also be specified depending on the application. If any one of the specified test values fails to meet its tolerance requirements, the device is usually considered to have failed the calibration requirements.

12.3.9 Test Procedures

There are many methods that can be used to verify or calibrate measuring and test equipment. Some became industry standards due to widespread or common use, however, there are no sets of law that determine the methods to be used. There are accepted practices for certain types of measurements (industry standards) and deviation from these methods might invoke question and/or suspicion from customers or auditors. In general, any logical, reproducible, documented process that uses proven standards is deemed to be acceptable. These documents can be in the form of published industry practices, books, manufacturer's manuals, and internally or externally produced layouts. Software programs, on-line procedures, and hard copy media are acceptable.

Typical calibration procedures include information on calibration intervals, the standards to be used, the setup, the methodology, the measurements to be taken, the allowable tolerances, the data to be recorded, and corrective action such as repair and adjustment procedures. As an example, the essential aspects of a documented calibration procedure are given in Figure 12.4.



QUALITY CORPORATION

CALIBRATION
OF
ACME 545 FLOW METERS

Version

Issue date 5-10-97

pg. 1 of 6

Interval:
Calibrate once per year (June)

Standards:
ACME 545 flow meter standard

pg. 2 of 6

Setup:

Configure the standard and test device as shown below, make sure all connections are secure

545 Flow meter | 545 Standard

Flow direction →

Procedure:

Once setup is complete,

- Take measurements
- Record measurement values on data sheet
- Compare to listed tolerances
- If values exceed tolerance refer to "Adjustments" on pg 6

pg. 4 of 6

Data Sheet

Nom	Tol	Act
0	+/-.1	0
5	+/-.2	5.001
10	+/-.3	10.01
20	+/-.4	19.99
50	+/-.5	49.98

Signature_______Date________

pg. 5 of 6

Adjustments:

Turn knob A to position B, dial in offset, return to original position. Retest. If values are not within tolerance, return unit to ACME for replacement

pg. 6 of 6

Figure 12.4 A documented calibration procedure.

12.3.10 Reliability and Performance Metrics

Measurement data can be used to determine various characteristics about equipment, methods and processes, as well as test acceptance. Measurement reliability is the probability that all measurement attributes of a piece of equipment are in conformance with performance specifications and will be so for a specified period. Systematic errors tend to increase with time, which leads to the growth of measurement uncertainty. To set a 95% reliability target means that test intervals have been defined so that 95% of equipment tested on that interval is found to be within tolerance. Reliability targets are used to determine calibration intervals that coincide with cost and quality requirements.

Measurement variability is commonly expressed in terms of stability, bias, linearity, repeatability, and reproducibility (see Section 12.3.12). The terms stability, bias and linearity are used to describe the location of measurement values in relation to stated values. The terms repeatability and reproducibility are used to describe the width or spread of measurements in relation to each other. Measurement variance is the basis for quantifying the overall performance characteristics of measurement systems and allowing for comparisons between test methods and equipment. Measurement reliability and variability metrics can be used to determine the most efficient process available for testing or obtaining a measurement.

12.3.11 National Accreditation Board of Approved Laboratories (NABAL)

The Government of India has started a national programme entitled National Accreditation Board of Approved Laboratories (NABAL). The Government of India has also constituted an Experts Panel on Electronic Measurements, under the chairmanship of Director General, Standardization, Testing, Quality and Certification (STQC), for this purpose. The Experts Panel has finalized the general/specific requirement for approval of electronic calibration laboratories covering various parameters, ranges, measurement uncertainties, and accuracies from dc to microwave frequency ranges, environmental conditions, namely, calibration temperature, humidity, dust, light, moisture, magnetic field, shock, EMI/RFI levels, power supply requirements, repair and maintenance, etc. Expert panel has also drawn a panel of assessors of eminent scientists/engineers to assess the calibration laboratory from a system's angles and uniformity. Mock exercise has also been carried out to see the suitability of specific criteria and preparedness of the laboratory seeking approval. This programme was formally launched in 1995. Based on the input of Expert Panel, Department of Science & Technology (DST) has published a document entitled "Guidelines, Criteria and Application Performa for Accreditation of Calibration Laboratories in the field of Electronics Measurements". This document is available from DST.

12.3.12 Definitions of Some Commonly Used Terms

- *Accuracy.* This is the closeness of the agreement between the result of a measurement and the value of the measurand.
- *Bias.* The extent to which the average of a long series of repeat measurements made by an instrument on a single unit differs from the true average.
- *Calibration.* The set of activities to ensure that measuring instruments maintain their metrological characteristics. The instrument is compared to a standard, and adjustments are made if necessary. A calibration program includes inspection of the measuring instrument, maintaining a schedule, and keeping records.
- *Calibration interval.* This is the time between calibrations for each measuring instrument.
- *Confidence level.* The probability that a measured value will be within the tolerance limits. The confidence level for standards is usually set at 95%.
- *Linearity.* A measure of the degree to which a response maintains a linear relationship with given input stresses.
- *Measurand.* A quantity subjected to measurement or the characteristic that is measured.
- *Measurement and test equipment (M&TE).* An instrument, device or system of devices used to make measurements. The standards used to verify measuring and test equipment are also classified as measuring and test equipment.

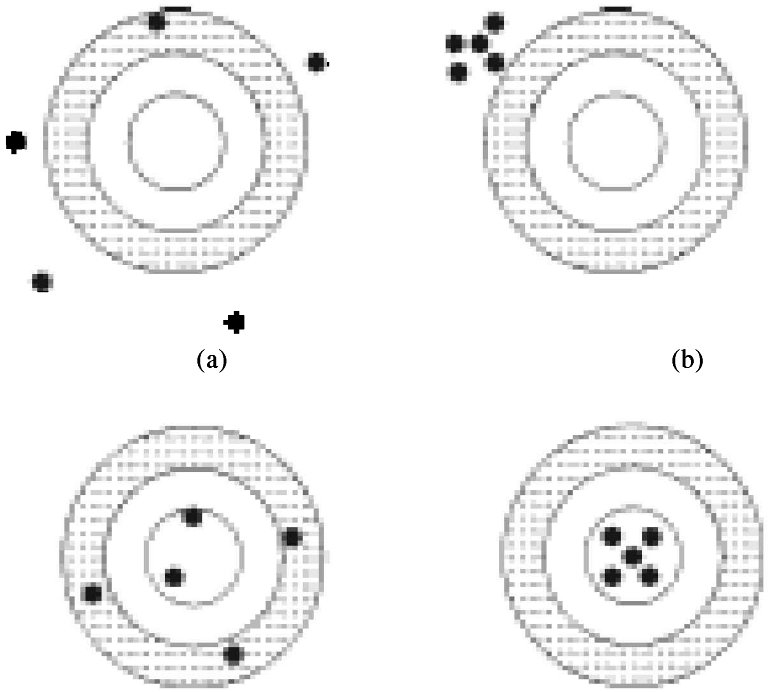
- *Measurement standard.* A material measure, instrument, system or reference intended to define or reproduce known values of a quantity. Standards are used to check the accuracy and precision of measuring instruments and lower level standards.
- *Measurement.* The value of a measurand.
- *Nominal value.* This is the stated or par value of a measurement.
- *Precision.* The extent to which an instrument gives constant results when making repeat measurements on the same unit. That is the closeness of agreement between independent test results obtained under stipulated conditions
- *Primary standards.* These are the highest level standards and are usually maintained by an internationally recognized standards organization. All other standards must have traceability to the primary standards. In the United States, NIST maintains many primary standards, although some private companies may maintain their own primary standards.
- *Resolution.* Resolution is just the discrimination that the instrument can show. It is the smallest change in the quantity being measured.
- *Repeatability.* This is the closeness of the agreement between the results of successive measurements of the same measurand carried out under the same conditions of measurement. That is the ability to reproduce the same level of accuracy under the same conditions. The test is carried out using the same measurement procedure, the same measuring and test equipment, and the same environmental conditions.
- *Reproducibility.* This is the ability to reproduce the same level of accuracy under any given condition. Changed conditions may include principle or method of measurement, observer, measuring and test equipment, location, conditions of use and time.
- *Stability.* This is the ability of a measuring instrument to maintain constant its metrological characteristics over a specified time interval.
- *Tolerance limits.* These are the measurement specifications that bound the acceptance region for a measured characteristic.
- *Traceability.* This is the evidence that working standards or transfer standards can be traced to primary standards through an unbroken chain of comparisons.
- *Transfer standards.* These are secondary standards that can be traced to primary standards. Although there is no limit on how many transfer standards can be between the working and primary standards, the number should be as small as possible.
- *Uncertainty.* This is the error associated with a standard together with the confidence level. An example of an uncertainty statement: There is a 95% confidence level that the stated value will be within ±0.0001 of the true value.
- *Working standards.* These are standards actually used to test the measuring equipment. Working standards are usually calibrated with transfer standards although they can be calibrated with primary standards.

Difference between accurate and precise

We will use some graphics to illustrate this. Let us have some archery or shooting targets. Five marksmen are aiming for the centre "bulls-eye". This is analogous to making a perfect measurement with the "bull" being the conventional "true value". So aim is taken and five rounds are fired. The results are shown in Figure 12.5. The shots of Figure 12.5(a) are widely distributed and mostly off-target—this person is obviously a beginner, both inaccurate and unrepeatable. The shots of Figure 12.5(b) are closely grouped but they have all missed the target completely. This person is precise but inaccurate. The third person has reliably hit the target (Figure 12.5(c)) but the shots are dispersed—so we have accuracy (two in the "bull") but imprecision. The final target [Figure 12.5(d)] shows the way it should be done—little deviation from "true" every time, showing both accuracy and precision.

As far as calibration is concerned, the attribute *accurate* often also implies *precise* but it is worth remembering that it may not be the case. Conversely, the person who claims that his product is precise may not be making any claim at all for its correctness (relationship to national standards).

The degree of accuracy and precision results from the combined effect of measuring equipment, technique, environmental conditions and the characteristics of the item being tested. If a series of repeated measurements were made and the data plotted as a histogram (bar graph), the shape described by the bar-heights would represent the distribution.



(a) (b)

(b) (c)

Figure 12.5 Difference between accuracy and precision.

12.4 ELECTRICAL STANDARDS

All electrical quantities are based upon three basic parameters: voltage, current, and resistance. These are further related with each other through Ohm's law. So if two parameters are known, the third can be calculated. National laboratories, nowadays, maintain voltage and resistance standards and base the values of the other standards on these two.

12.4.1 Voltage Standard

The Josephson array voltage standard is used worldwide as the primary voltage standard in all major calibration laboratories.

A Josephson junction is made by sandwiching a thin layer of a nonsuperconducting material between two layers of a superconducting material. The devices are named after Brian Josephson, who predicted in 1962 that pairs of superconducting electrons (Cooper Pairs) could *tunnel* right through the nonsuperconducting barrier from one superconductor to another. He also predicted the exact form of the current and voltage relations for the junction. This was later verified experimentally and Josephson was awarded the 1973 Nobel Prize in Physics for his work.

To understand the unique and important features of Josephson junctions, it is first necessary to understand the basic concepts and features of superconductivity. If we cool some metals and alloys to very low temperatures (within 20 degrees or less of absolute zero), a phase transition occurs. At this *critical temperature*, the metal goes from what is known as the normal state, where it has electrical resistance, to the superconducting state, where there is essentially no resistance to the flow of direct electrical current. The newer high-temperature superconductors, which are made from ceramic materials, exhibit the same behaviour but at somewhat high temperatures.

The electrons in the metal become paired. Above the critical temperature, the net interaction between two electrons is repulsive. Below the critical temperature, the overall interaction between two electrons becomes very slightly attractive as a result of the electrons' interaction with the ionic lattice of the metal. This very slight attraction lets electrons to drop into a lower energy state, opening up an energy *gap*. Because of the energy gap and the lower energy state, electrons can move (and therefore current can flow) without being scattered by the ions of the lattice. When the ions scatter electrons, it causes electrical resistance in metals. There is no electrical resistance in a superconductor, and therefore no energy loss. There is, however, a maximum supercurrent that can flow, called the critical current. Above this critical current the material is normal. There is one other very important property: when a metal goes into the superconducting state, it expels all magnetic fields, as long as the magnetic fields are not too large.

In a Josephson junction, the nonsuperconducting barrier separating the two superconductors must be very thin. If the barrier is an insulator, it has to be of the order of 30 Å thick or less. If the barrier is another metal (nonsuperconducting), it can be as much as several microns thick. Until a critical current is reached, a supercurrent can flow across the barrier; electron pairs can tunnel across the barrier without any resistance. But when the critical current is exceeded, another voltage will develop across the junction. That voltage will depend on time—that is, it is an ac voltage. This in turn causes a lowering of the junction's critical current, causing even more normal current to flow—and a larger ac voltage.

The frequency of this ac voltage is nearly 500 gigahertz (GHz) per millivolt across the junction. So, as long as the current through the junction is less than the critical current, the voltage is zero. This is called dc Josephson effect. As soon as the current exceeds the critical current, the voltage is not zero but oscillates in time (ac Josephson effect). Detecting and measuring the change from one state to the other is at the heart of the many applications for Josephson junctions.

The Josephson effect is a quantum mechanical phenomenon, which generates a number of constant voltage steps in the *I-V* characteristic of a Josephson junction (Figure 13.6) at the voltages

$$V_n = \frac{hnf}{2e}$$

where n is an integer which depends upon the biasing current of the junction, f is the frequency of an external microwave source, h is Planck's constant and e the electron charge. $(2e/h)$ is also called the Josephson constant, $K_{J\text{-}90}$ (90 indicates the year 1990) and has the value 483597.9 GHz/V.

For $n = 1$ and $f = 71$ GHz, junction voltage = 146.8 μV and for $n = 6930$ and same frequency, junction voltage = 1.018 V.

Depending on the technology used, the critical current of the first voltage step is typically about 1 mA with a current width of about 100 μA.

The standard volt is now defined as the voltage required to produce a Josephson frequency of 483,597.9 GHz.

In practical experiments, the fundamental limits are set by the reproducibility of the frequency of the external microwave source, which is only one part in 10^{12} when modern caesium clocks are used to determine the drive frequency, f. The frequency is usually chosen to about 70 GHz, which gives constant voltage steps as multiples of about 145 μV. The maximum voltage across a single junction is about 1 mV ($n = 7$). By connecting a large number of junctions in series (Josephson array), voltages up to more than 10 V have been generated. Josephson junction standards can yield voltages with accuracies of one part in 10^{10}.

To obtain a rapid programmable dc voltage standard with the possibility of synthesizing low frequency ac waveforms, several solutions have been studied. One method is to connect the junctions in a binary configuration and allow each segment to be biased independently to the $n = 0$ or $n = \pm 1$ quantum state. By controlling the biasing of the different segments with fast switching current sources controlled from a system computer, n can be changed and any combination of voltages in multiples of the voltage across a single junction can be obtained. Table 12.2 gives typical Josephson junction array chip specifications made from niobium superconductor.

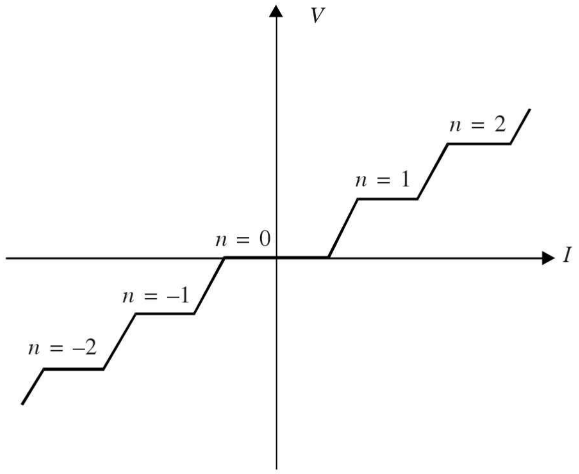


Figure 12.6 Current-voltage characteristic of a typical Josephson junction.

Table 12.2 Typical Josephson junction array chip specifications

Parameter	10 V *array chip*	1 V *array chip*
Chip size	25 by 11 mm^2	19 by 5 mm^2
Number of junctions	13,920 JJs	2,400 JJs
Operating range	–11 V to +11 V	–2.5 V to +2.5 V
Technology	refractory all niobium full wafer process	refractory all niobium full wafer process
Operating temperature	4.2 K	4.2 K
Operating frequency	70–75 GHz	70–75 GHz
Lead and bond wire resistance	< 3 ohm	< 3 ohm

Pure metals that have superconducting property are called Type I superconductors. The identifying characteristics are zero electrical resistivity below a critical temperature, zero internal magnetic field, and a critical magnetic field above which superconductivity ceases. Their critical temperature varies from 0 to 9.46 K and critical field from 10^{-1} to 10^{-5} tesla.

Superconductors made from alloys are called Type II superconductors. Besides being mechanically harder than Type I superconductors, they exhibit much higher critical magnetic fields. Type II superconductors such as niobium-titanium (NbTi) are used in the construction of high field superconducting magnets. Their critical temperature varies from 10 to 23.2 K and critical field from 1.5 to 44 tesla.

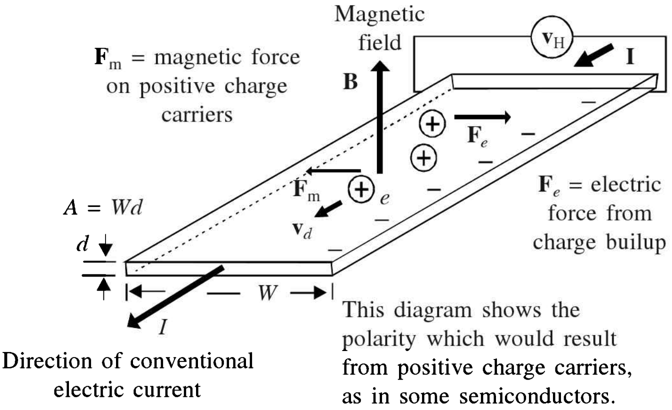
Voltage transfer standards

A battery-operated voltage measuring system can be generally used to calibrate 1.018 V output of temperature controlled zener diode standards with an estimated uncertainty of about 2×10^{-8}. The zener diode is used as a transfer standard for calibrating standard cell and other instruments.

12.4.2 Resistance Standard

Since 1 January 1990, the Quantum Hall effect (QHE) has been used by most national metrology institutes as the primary resistance standard.

If an electric current flows through a conductor in a magnetic field, the magnetic field exerts a transverse force on the moving charge carriers which tends to push them to one side of the conductor. This is most evident in a thin flat conductor as illustrated in Figure 12.7. A buildup of charge at the sides of the conductors will balance this magnetic influence, producing a measurable voltage between the two sides of the conductor. The presence of this measurable transverse voltage is called the Hall Effect after E.H. Hall who discovered it in 1879.


Magnetic field
B
v_H
I
F_m = magnetic force on positive charge carriers
F_e
F_m
e
v_d
A = Wd
d
W
I
F_e = electric force from charge builup
Direction of conventional electric current

This diagram shows the polarity which would result from positive charge carriers, as in some semiconductors.

Figure 12.7 Principle of Hall Effect.

The magnetic force is $\mathbf{F}_m = e\mathbf{v}_d\mathbf{B}$ where $\mathbf{v}_d$ is the drift velocity of the charge.

The current expressed in terms of the drift velocity is

$$\mathbf{I} = neA\mathbf{v}_d$$

where n is the density of charge carriers and e the electron charge. Then

$$\mathbf{F}_m = \frac{e\mathbf{IB}}{neA}$$

At equilibrium

$$\mathbf{F}_m = \mathbf{F}_e = \frac{\mathbf{v}_H e}{W}$$

Equating the two equations, we get

$$\mathbf{v}_H = \frac{\mathbf{IB}}{ned}$$

Hall resistance, R_H can then be defined as

$$R_H = \frac{B}{ned}$$

R_H can thus be measured to find n the density of carriers in the material. Often this transverse voltage is measured at fixed current and the Hall resistance recorded. It can easily be seen that this Hall resistance increases linearly with magnetic field.

Quantum Hall effect

Hall performed his experiments at room temperature and with moderate magnetic fields of less than one tesla (T). At the end of the 1970s, researchers used extremely low temperatures (only a few degrees from absolute zero, i.e. around –272°C) and very powerful magnetic fields (max approx. 30 T). They studied the Hall effect in the type of semiconductor design used in the electronics industry for manufacturing low-noise transistors. The material contains electrons which, though trapped close to an internal surface separating two distinct parts of the material, are highly mobile along the surface.

In such a layer at low temperatures, electrons can be caused to move as if on a plane surface, i.e. in two dimensions only. This geometrical limitation leads to many unexpected effects. One is that the Hall Effect changes character. This is seen most simply when one measures how the Hall resistance varies with the strength of the applied magnetic field.

In 1980 the German physicist Klaus von Klitzing discovered in a similar experiment that the Hall resistance does not vary in linear fashion, but *stepwise* with the strength of the magnetic field. The steps occur at resistance values that do not depend on the properties of the material but are given by a combination of fundamental physical constants divided by an integer. We say that the resistance is quantized. At quantized Hall resistance values, normal ohmic resistance disappears and the material becomes in a sense superconducting.

Thus, the Quantum Hall Effect (QHE) is observed at very low temperatures (<1 K) in a two-dimensional electron gas of high mobility ($\mu > 2T^{-1}$) which has been placed in a strong magnetic field perpendicular to the plane of conduction. A two-dimensional electron gas can, for example, be produced in a MOS field effect transistor or in a GaAs/AlGaAs heterostructure. If, under these experimental conditions, the Hall voltage is measured as a function of the magnetic flux density B, then plateaux with constant Hall voltage are observed. On these plateaux the Hall resistance R_H is quantized and is equal to

$$R_H = \frac{h}{i \cdot e^2}$$

where i is an integer. This is called *Integer Quantum Hall Effect* (IQHE). The quantum resistance h/e^2 is also known as the von Klitzing constant R_K. The longitudinal voltage along the probe displays a markedly oscillatory behaviour (Shubnikov-de-Haas Effect). The plateaux in the Hall voltage thereby fall together with extended minima in the longitudinal voltage. As the temperature is lowered, the voltage in these minima becomes so small as to be unmeasurable, and consequently as absolute zero is approached current flow through the probe shows zero dissipation.

Figure 12.8 shows the IQHE in a GaAs-GaAlAs heterojunction, recorded at 30 mK. The QHE can be seen at liquid helium temperatures, but in the millikelvin regime the plateaus are much wider. Also included is the diagonal component of resistivity, which shows regions of zero resistance corresponding to each QHE plateau. In this figure the plateau index is, from top right, 1, 2, 3, 4, 6, 8.... Important points to note are:

- The value of resistance only depends on the fundamental constants of physics: e the electric charge and h Planck's constant.
- It is accurate to 1 part in 100,000,000.
- The QHE can be used as a primary resistance standard, although 1 klitzing is a little large at 25812.807 ohm (for $i = 4$). The relative uncertainty of this value within the SI is 2×10^{-7}.

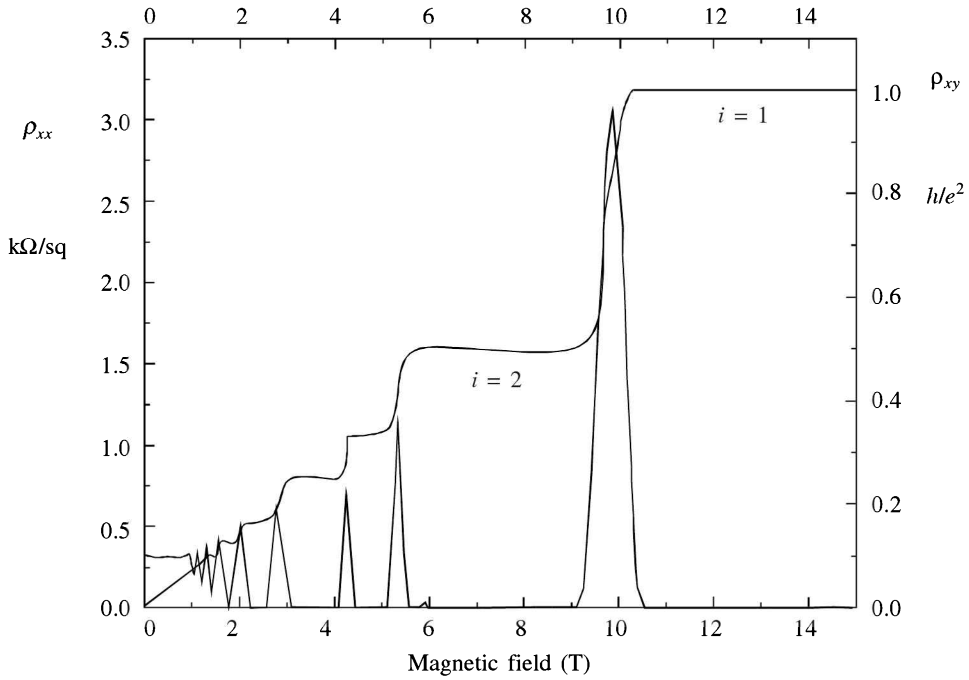


Figure 12.8 Integer Quantum Hall Effect.

In high mobility semiconductor heterojunctions the plateaux are much narrower than for lower mobility samples. Between these narrow IQHE, more plateaux are seen at fractional values of i.

In their refined experimental studies of the QHE, using among other things lower temperatures and more powerful magnetic fields, Störmer, Tsui and their co-workers found to their great surprise a new step in the Hall resistance which was three times higher than von

Klitzing's highest. They subsequently found more and more new steps, both above and between the integers. All the new step heights can be expressed with the same constant as earlier but now divided by different fractions. For this reason the new discovery was named the *Fractional Quantum Hall Effect* (FQHE).

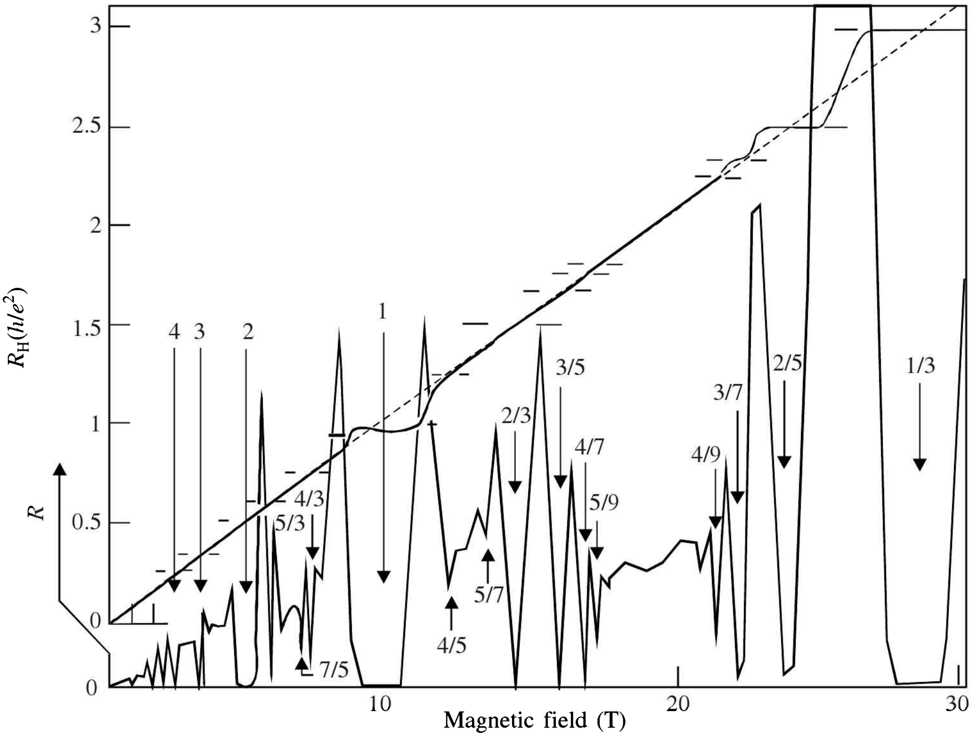


Figure 12.9 Fraction Quantum Hall effect.

In 1998 the Nobel Prize in Physics was awarded to Dan Tsui and Horst Stormer, the experimentalists who first observed the FQHE, jointly with Robert Laughlin, who succeeded in explaining the result in terms of new quantum states of matter.

We observe that a large number of precision resistance values can be obtained and used as primary standard for resistance.

Resistance transfer standards

Materials like manganin wire, glass-ceramic foil, precision metal film, air, etc. are used for fabricating resistors for transfer standards. Some resistance standards can be used in air and others in liquid medium (oil, kerosene). Some are available as single resistors, and many in the form of resistance boxes.

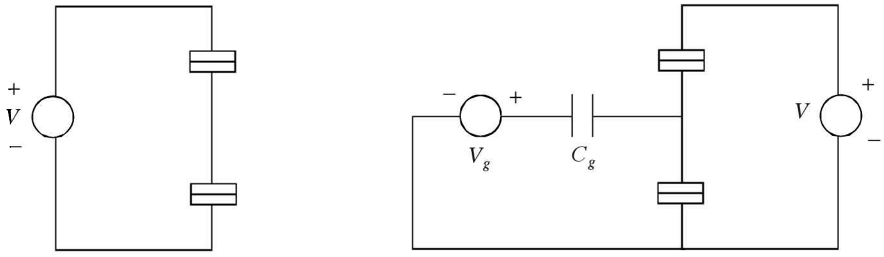
Using a high-precision resistance bridge, the traditional resistance standards are compared with the quantized Hall resistance, allowing them to be calibrated absolutely. These resistance standards serve in their turn as transfer standards for the calibration of customer standards. An accuracy of 1 × 10^{-9} can be obtained.

12.4.3 Current Standard

From a practical point of view, there is no pressing need for a fundamental current standard because we already have excellent fundamental standards of voltage (Josephson Effect) and resistance (Quantun Hall Effect). However, if we could get current from a device, then it can be combined with voltage from a Josephson effect device to create a resistance that can be compared to the quantum Hall resistance.

Single-electron tunneling (SET)

Consider two tunnel junctions defining a metallic island of total capacitance C, as shown in Figure 12.10(a). The basic condition needed for the observation and exploitation of SET effects is that the system must *notice* when the number of electrons on the island changes by one, despite the fact that there may be millions of electrons on the island. This condition is fulfilled when two requirements are met. First, the electrostatic energy needed to place a single extra electron onto the island, $e^2/2C$, must be large compared to the energy of thermal fluctuations, kBT. Second, the tunnel barriers must be sufficiently opaque so that the wavefunction of an extra electron on the island is well localized there. In other words, the lifetime associated with tunneling must be sufficiently long that the corresponding uncertainty in energy is small compared to the single-electron charging energy $e^2/2C$. This requirement is met when the tunnel resistance R is large compared to h/e^2. In a device that satisfies these two requirements, the number of excess electrons on the island is an integer and there is a significant energy barrier for changing this number. If a voltage is applied across the two junctions in Figure 12.10(a), current will not flow through the system unless the voltage source can supply enough energy to overcome this energy barrier. This phenomenon, which is readily observed in real systems, is known as the Coulomb blockade of electron tunneling.



(a) (b)

Figure 12.10 (a) A single island of total capacitance C defined by two tunnel junctions (double box symbols), and (b) A single-electron tunneling transistor created by adding a gate capacitor and voltage source.

To make useful devices based on the ideas just described, there must be a way of manipulating the charging energy barrier so that tunneling can be turned on and off at will. This is accomplished by adding a third electrode to the system, as shown in Figure 12.10(b). A voltage applied to this electrode polarizes the island with a charge that can be varied continuously. When this polarization charge is $e/2$, the energy of the system with 0 and 1 extra electrons on the island is the same, and the system does not *notice* if one extra electron tunnels on or off the island, i.e., the charging energy barrier has been reduced to zero. The third electrode that allows the energy barrier to be manipulated in this way is called a gate electrode, and the complete three-terminal device in Figure 12.10(b) is called a single-electron tunneling transistor (SETT).

Controlled transfer of individual charges can be accomplished by putting several junctions and islands in series, along with a gate electrode for each island, to form a device called an *electron pump.*

Single-electron tunneling (SET) devices provide a means of manipulating individual electrons and detecting the motion of these electrons with extraordinary precision. Their potential impact on metrology and fundamental constants was recognized early in the development of the field during the 1980s. By the early 1990s, several SET devices had demonstrated the ability to detect charges much smaller than e and to transfer individual charges from one electrode to another. In the last few years, the performance of these devices has been pushed to the levels needed for fundamental standards and high-precision measurements: SET electrometers can detect $\sim 10^{-5}e$ in a 1 Hz bandwidth; electron traps can store individual charges for hours; electron pumps can transfer hundreds of millions of individual electrons with an uncertainty of $\sim 10^{-9}$. Fundamental standards based on these capabilities are being pursued at several national measurement institutes throughout the world, and SET devices are at the threshold of making important contributions to practical and fundamental metrology.

Of the various potential applications for SET standards of current and capacitance, the one that best illustrates the impact of manipulating individual electrons on metrology and fundamental constants is the realization of the quantum metrology triangle. The idea, first proposed by Likharev and Zorin in 1985, is illustrated in a simplified form in Figure 12.11. A frequency source traceable to an atomic clock is used to drive a Josephson voltage source that produces a voltage $V_J = f/K_J$ and an SET pump that produces a current $I_{SET} = Q_X f$, where theory predicts that $K_J = 2e/h$ and $Q_X = e$. The ratio V_J/I_{SET} defines a resistance that can be compared with the Quantum Hall resistance $R_H = R_K = h/e^2$. The triangle is essentially an application of Ohm's law, $V = IR$, to the quantities produced by the three quantum effects.

The uncertainty of most quantum metrology triangle experiments envisioned thus far is directly limited by the small SET current because the voltage produced by passing this current through a quantum Hall device is very small. The maximum current that can be produced with the required uncertainty is thus the primary criterion for comparing different SET current standards.

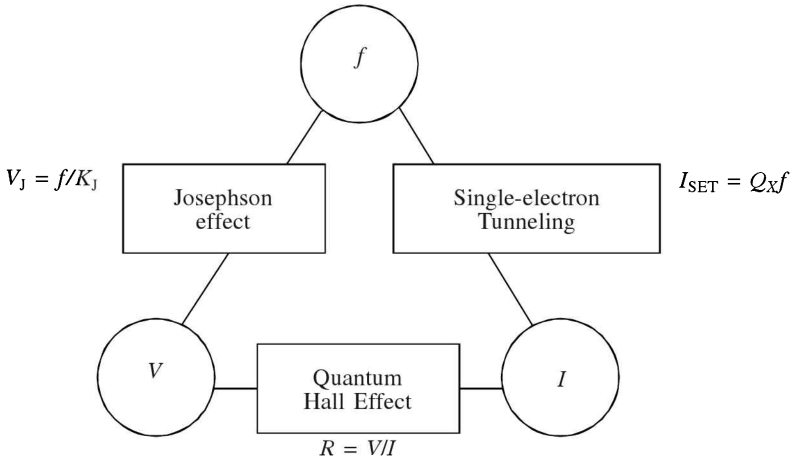


Figure 12.11 The quantum metrology triangle.

12.4.4 Capacitance Standard

The present representation of the SI farad is through silica-based artifact capacitors. Although these capacitors are of high quality, they are susceptible to drift in time and they may depend on other parameters such as temperature, pressure, and frequency. The metrology community, including both the national standards laboratories and domestic secondary calibration laboratories, needs a capacitance representation that is based on fundamental physical principles and not on properties of individual physical artifacts.

NIST researchers announced in the Sept. 10, 1999, issue of Science magazine that they have developed the first working prototype of a new standard for measuring the storage of electrical energy—known as capacitance—based on a means of manipulating and counting electrons one at a time (Single Electron Tunnling).

The new standard is the latest to rely on a fundamental property of nature and not be determined by a constructed artifact, classical physics or a combination of both. Nature-based standards currently in use include length (the metre) and time (the second), both defined by precise measurements of the vibrations of the cesium atom.

The work on the capacitance standard depends on two technologies developed at NIST. The first is an electron pump, based on ultra-small electrical devices called tunnel junctions. Operating at temperatures less than one-tenth of a degree above absolute zero, the pump passes and counts individual electrons with an uncertainty of 0.01 parts per million. In other words, the pump would miss tallying only one electron in every 100 million passing through.

The second NIST-developed technology is a cryogenic (very low temperature), vacuum-gap capacitor. The capacitor is designed so that very few of the electrons placed upon it *leak out*, dramatically improving the accuracy of the capacitance measurement. In addition, it does not contain the dielectric materials (which are good but imperfect electrical insulators) that lessen the effectiveness of ordinary capacitors.

The capacitance measurement is made after the cryogenic pump places about 100 million electrons on the capacitor. The resulting voltage generated across the capacitor is determined, and the capacitance is then calculated as the ratio of pumped charge to measured voltage.

In defining capacitance, we consider a transfer of charge between two conductors. The transferred charge Q causes a potential difference ΔV between the conductors, and the capacitance is simply $C = Q/\Delta V$. Since nature provides abundant electrons having identical quantized charges, it is natural to use them as the basic unit of charge in implementing the definition of capacitance. This is precisely what is done in the SET capacitance standard. A SET pump is used to place N electrons onto one electrode of a cryogenic, three-terminal, vacuum-gap capacitor, then the resulting ΔV is measured, and the capacitance is given by $C = Ne/\Delta V$. The same cryogenic capacitor can also be measured in a conventional bridge configuration so that its value can be transferred to an artifact at room temperature for convenient comparison with other capacitors.

NIST's present primary capacitance standard has been measured to an uncertainty level of 0.01 parts per million. Although very accurate, the measurement requires complex calculations, takes months to complete and is best done at a national measurement laboratory such as NIST. The system for this new standard is simpler to use, makes measurements faster and can be set up at any site with the necessary equipment.

Capacitance transfer standards

Dielectrics like air, silvered-mica, polystyrene, polycarbonate, fused silica, etc. are used for fabricating capacitors for transfer standards. Some are available as single capacitors and others in the form of decade boxes.

12.4.5 Inductance Standard

Inductance can be derived from ohm and farad. See Maxwell and Hay bridge in Chapter 5. A wide range of transfer standards are available as fixed and variable for various values.

12.5 TIME AND FREQUENCY STANDARDS

No watch or clock is completely accurate. Each has its own errors due to rate imperfections and errors in setting. The correct time is simply based on an agreed standard. Currently, Universal Time Coordinated (UTC) has been established as the world time scale.

No physical clock keeps UTC. UTC is established by the Bureau International des Poids et Mesures (BIPM) based on an aggregate of data from timing laboratories throughout the world, and from input from the International Earth Rotation Service (IERS).

Practical, precise timing came with the invention of the quartz-crystal oscillator and quartz-crystal filters, which are essential elements for radio, radar and television with their enormous, far-reaching impact on our society. Information is literally flowing at the speed of light in computers and communications systems. With their ever-increasing capabilities, they would not work without precise timing of gates and network nodes. The arrival of atomic clocks provided even more accuracy and opened new vistas. Unprecedented navigation is now literally at our fingertips using relatively inexpensive hand-held GPS (Global Positioning System) receivers; yet the heart of GPS is an atomic-clock synchronized system. Since their invention, the accuracy of atomic clocks has improved, on the average, by a factor of two every two years. The current best accuracy for the determination of the second results in a time error of ±0.3 nanoseconds (billionths of a second) per day. This is equivalent to ±1 second in 10 million years.

Because time and frequency can be measured so accurately, time and frequency devices are often used to measure other fundamental quantities, such as the volt, the ampere, the ohm, and the metre. For example, the General Conference of Weights and Measures (CGPM, Conférence Générale des Poids et Mesures) redefined the metre as "the length of the path travelled by light in vacuum during a time interval of 1/299,792,458 of a second (17th CGPM, 1983, Resolution 1)." It is expected that eventually, all base units in metrology for the support of technological development will be traceable back to the second.

The Global Positioning System (GPS) is a classic example of using precise timing for accurate positioning. Many navigation systems predating GPS also use atomic clocks, but they are ground-based. GPS features a set of 24 orbiting satellites, each with a synchronized atomic clock on board. It effectively puts a super-accurate clock in the sky for everyone to see with the eyes of modern technology. At any given time and at any point on Earth, at least four of these satellites can be seen. A precise timing device in a GPS receiver is used by its computer to calculate the time of flight of the signal from each of the observable satellites. Since the signals travel at the velocity of light, and this is known exactly, the receiver's computer can turn the time of flight into a very accurate estimate of the distance to each satellite—accounting for some delays in both the neutral and ionized parts of the atmosphere. GPS satellites also broadcast their positions. Information received from four satellites yields four equations which can be solved for four unknowns: latitude, longitude, altitude, and GPS-system time. In this process, the high accuracy of the GPS atomic clocks is transferred to the precision quartz-crystal clock inside the receiver. Therefore, high accuracy position and time are readily available (from which velocity can be deduced), and a multitude of users are capitalizing on this. GPS is like a free utility with application opportunities limited only by our imaginations. As a result, the number of users, and the variety of uses of GPS have literally exploded.

Measuring and comparing time and frequency of clocks remote from each other can be accomplished in a wide variety of ways and with a variety of accuracies and stabilities. Usually, the better the accuracy and stability, the more expensive will be the system. During the last two decades with the advent of the Global Positioning System, great progress has been made in transporting time and frequency. With the atomic clocks on board the satellites, GPS acts like a portable clock in the sky, continuously available anywhere on Earth via very inexpensive receivers. Accuracies better than one microsecond (one millionth of a second) with respect to UTC are readily available by the proper use of GPS timing receivers. Satellite techniques, in general, have demonstrated significant advantages over terrestrial techniques, such as improvements in accuracy, integrity, availability, continuity, coverage of service, and perhaps more importantly, cost. However, the performance one sees from a clock is affected by signal propagation variations and delays, changes intrinsic to the clock, and environmental perturbations on the clock.

Definition of the second and its general importance

In 1967, it was agreed by the 13th General Conference of Weights and Measures (CGPM [Conférence Générale des Poids et Mesures] Resolution 1) that: "The second is the duration of 9,192,631,770 periods of the radiation corresponding to the transition between the two hyperfine levels of the ground state of the cesium-133 atom." The current best accuracy for the realization of the second so defined is equivalent to ±1 second in 10 million years. At the time of the definition, the atomic second was made to agree as much as possible with the ephemeris second based on astronomical measurements. The 9,192,631,770 Hz assigned to the above cesium transition was the result of a three-year cooperative between L. Essen at the National Physical Laboratory (NPL) in Teddington, England and W. Markowitz at the United States Naval Observatory (USNO) in Washington DC NPL supplied the cesium clock, and USNO provided the astronomical measurements via a clever dual-rate moon camera which could simultaneously compensate for the different movement of the moon and the stars. The occultation of the latter by the former gave a very accurate estimate of ephemeris time.

A very accurate frequency has three principal advantages over the other base standards in metrology: first, no other standard can presently be measured so accurately; second, high levels of accuracy and stability can be obtained relatively inexpensively; and third, it can be communicated via electromagnetic waves (as in radio transmission, microwave transmission and laser transmission). With the development of these time and frequency techniques, it is not surprising that there has been a dramatic increase in the use of precise timing over the last few decades, and this trend is expected to continue. GPS is a classic example.

How to access UTC?

Since UTC is not directly available as a clock, real-time approximations to it are made available from 50 timing centres around the world. In India, National Physical Laboratory, New Delhi is maintaining this time. A good percentage of the timing centres keep within 100 nanoseconds of UTC, and a few of them are usually within 10 ns of UTC.

QUESTIONS

12.1 Explain the meanings of calibration and traceability. Why are these necessary?

12.2 What are the ISO 9001 and ISO 9002 requirements for caliberation?

12.3 What are the basic elements of a calibration programme?

12.4 Explain the hierarchy of traceability chain.

12.5 Show the various echelons for the traceability of accuracies in electronics measurement under the national calibration policy.

12.6 The following instruments require calibration. Where should they be calibrated and at what accuracy?

(a) Digital multimeter of 100 ppm accuracy being used by Electronics Regional Test Laboratory, New Delhi.

(b) Standard resistor of ±1% accuracy being used by B.I.T.S., Pilani.

(c) Fluke Voltage Calibrator of 10 ppm accuracy being used by Defence Research Laboratory, Hyderabad.

(d) K Type thermocouple of ±1% accuracy being used by Steel Plant, Salem.

12.7 Draw the traceability diagram of your Instrumentation Lab. to the National Physical Laboratory, New Delhi.

12.8 For the calibration of a 0.1% accurate digital multimeter at the shop-floor level, what are the calibration protocols required at various echelons? Show them through a diagram.

12.9 What is meant by measurement uncertainty? Classify the components of uncertainty.

12.10 How do you compute the instrument capability?

12.11 (a) Define the following:

- Primary Standards
- Transfer Standards
- Working Standards

(b) Explain graphically the difference between accuracy and precision.

12.12 How are voltage primary and transfer standards realized?

12.13 Explain Josephson Array Voltage Standard.

12.14 How are resistance primary and transfer standards realized?

12.15 Explain Quantum Hall Effect.

12.16 What is Single Electron Tunneling Effect? How is it used to generate current and capacitance standards?

12.17 Explain quantum metrology triangle.

12.18 Explain time and frequency standards.

Chapter 13

Standards in Quality Management

13.1 INTRODUCTION

Most organizations, industrial, commercial or governmental, produce a product or service intending to satisfy a user's needs or requirements. Such requirements are often incorporated in specifications. However, technical specifications may not in themselves guarantee that a customer's requirements will be consistently complied with. For example, there may be deficiencies in the specifications or in the organizational system to design and produce the product or service. Consequently, this has led to the development of quality system standards and guidelines that complement relevant product or service requirements given in the technical specifications.

Pressure for some sort of quality standards came initially from the military after the Second World War. The major companies in the automotive industry also began to develop their own quality system standards. The British Standards Institution (BSI) developed their military standards into the BS 5750 series (Part 1, 2 and 3: 1979). These have since been developed into the ISO 9000 series. ISO 9000 guidelines result in the following Deming's chain reaction:

- Quality improves
- Costs decrease because of less rework, fewer mistakes, fewer delays, better use of machine-time and materials
- Productivity improves
- Market share increases due to better quality and lower price

- Business gets established
- Jobs remain secure
- More jobs are provided

Each organization tends to adopt first ISO Quality System, then Total Quality Management (TQM) and finally World Class Quality.

In this chapter, we will study some important aspects of ISO Quality Management guidelines.

13.2 ISO 9000 QUALITY MANAGEMENT SYSTEM

The acronym-like term ISO stands for The International Organization for Standardization; it is based in Geneva, Switzerland. Founded in 1947 for the purpose of advancing standardization around the world, this non-governmental organization is now comprised of over 130 member countries. Each country is represented by its respective national standards body and participates in developing standards to facilitate trade of goods and services in the global marketplace. The standards developed apply not only to economic related activities but also cover the related technology and science involved in these pursuits.

The ISO 9000 series of quality management standards was developed by the ISO/TC 176 (ISO Technical Committee 176) convened in 1979. Five national associations participated in the ISO/TC 176 as conveners of subcommittees—AFNOR(Association Francaise de Normalization), ANSI (American National Standards Institute), BSI (British Standards Institute), NNI (Nederland's Normalisatie Institute), and SCC (Standards Council of Canada). Other member countries were also represented by their respective national standards bodies.

The committee set out to create a framework of the fundamental generic elements that would form the basis for a series of internationally recognized quality management standards, which it completed in 1982 and published in 1983. These underwent a major revision in 1994. The ISO 9000 series of standards represents the essential requirements that every enterprise needs to address to ensure consistent production and timely delivery of its goods and services to the marketplace. These requirements make up the standards that comprise the quality management system, and their generic nature allow for their application in any type of organization. Consistency in production and reliability in delivery are as important as what an organization is selling in today's marketplace. It is essential to consistently meet all of the customer's expectations all the time, every time, to keep them satisfied and loyal. If an organization does not do so, the marketplace, and with it the customers will take their business elsewhere. The ISO 9000 series is able to provide these quality management benefits to any organization of any size, public or private, without dictating how the organization is to be run. The system standards describe *what* requirements need to be met, *not how* they are to be met. This allows for diverse organizations to apply the same standards in a manner that reflects the reality of their business structure. In essence, this is like allowing each organization to meet the system requirements by implementing the standards in a manner that suits its own unique needs. Increasingly, certification to an internationally recognized quality management standard like the one from ISO 9000 series is becoming an important part of distinguishing an organization from its competition. The ability to be audited by an independent, third party organization is the foundation of its acceptance worldwide. Companies may now objectively determine an organization's capacity to supply goods or services that meet the specified requirements of the customer. This can attract more business as customers, both new and old, can have increased confidence in the organization's ability to meet their expectations. Often, just having an ISO 9000 certification will automatically qualify an organization as a potential supplier to companies that demand strict conformance to specified requirements. These standards have now become accepted around the world as the benchmark for all-quality management systems.

Some of the equivalent National Standards are listed in Table13.1.

Table 13.1 National equivalents to ISO 9000:1994

International	*United States*	*European*	*British*	*Japan*	*Indian*
ISO 9000	ANSI/ASQC Q90	EN ISO 9000	BS EN ISO 9000	JIS Z 9000	IS 14000
ISO 9001	ANSI/ASQC Q91	EN ISO 9001	BS EN ISO 9001	JIS Z 9001	IS 14001
ISO 9002	ANSI/ASQC Q92	EN ISO 9002	BS EN ISO 9002	JIS Z 9002	IS 14002
ISO 9003	ANSI/ASQC Q93	EN ISO 9003	BS EN ISO 9003	JIS Z 9003	IS 14003
ISO 9004	ANSI/ASQC Q94	EN ISO 9004	BS EN ISO 9004	JIS Z 9004	IS 14004

In short, the purpose of ISO 9000 is to:

- Harmonize the large number of National/International standards.
- Install and maintain a Quality System to strengthen competitiveness.
- Become globally integrated.
- Avoid multi third-party inspections/cost overruns.
- Establish Quality Systems Standards in place of product standards.
- Establish transparency of Quality Systems through evidence/documentation

13.2.1 Architecture of the ISO 9000 Series of Standards

The ISO 9000 series consists of five documents—three core quality system documents which are models of quality assurance, namely 9001/Q91, 9002/Q92 and 9003/Q93, and two supporting guidelines documents, 9000/Q90 and 9004/Q94 (See Table 13.2). The title of each document clearly indicates its purpose.

Table 13.2 The ISO 9000 family of standards

Standard	*Title*	*Contents in brief*
Quality assurance models for assessment under contractual situations		
ISO 9001: 1994	Quality systems—Model for quality assurance in design, development, production, installation and servicing.	Specifies quality system requirements for use when conformance to specified requirements is to be assured during design, development, production, installation and servicing. It provides a quality assurance model for demonstration by a supplier of its capability and for assessment by external parties.
ISO 9002: 1994	Quality systems—Model for quality assurance in production, installation and servicing.	Specifies quality system requirements for use when conformance to specified requirements is to be assured during production, installation and servicing. It provides a quality assurance model for demonstration by a supplier of its capability and for assessment by external parties.
ISO 9003: 1994	Quality systems—Model for quality assurance in final inspection and test.	Specifies quality system requirements and provides a quality assurance model for use where a supplier's capability to detect and control the disposition of any product nonconformity during final inspection and test needs to be demonstrated.
Quality management and quality assurance guidelines		
ISO 9000-1: 1994	Quality management and quality assurance standards—Part 1: Guidelines for selection and use.	Provides guidance for the selection and use of the ISO 9000 family of international standards on quality management and assurance. Clarifies the principal quality-related concepts needed for effective understanding and application of the standards and the distinctions and interrelationships between them.

(*Contd.*)

Table 13.2 The ISO 9000 family of standards (*Contd.*)

Standard	*Title*	*Contents in brief*
ISO 9000-2: 1993	Quality management and quality assurance standards—Part 2: Generic guidelines for the application of ISO 9001, ISO 9002 and ISO 9003.	Provides guidelines to enable users to have improved consistency, precision, clarity and understanding when applying the requirements of ISO 9000 series of standards.
ISO 9000-3: 1991	Quality management and quality assurance standards—Part 3: Guidelines for the application of ISO 9001 to the development, supply and maintenance of software.	Provides guidelines for software developers, maintainers and purchasers in the specification and implementation of ISO 9001 quality system requirements as applied to software.
ISO 9000-4: 1993	Quality management and quality assurance standard's—Part 4: Guide to dependability programme management.	Provides guidelines for implementing a quality system in an organization which provides services. Management is responsible for a quality policy and objectives, for ensuring to define responsibilities, authorities and lines of communication and for management review. It illustrates how the management must provide adequate resources to implement the quality system and to achieve the quality objectives including trained personnel and material resources.
ISO 10005: 1995	Quality management—Guidelines for quality plans.	Provides guidelines to relate generic requirements on quality system elements to the specific requirements of a particular product, project or contract. It also includes simplified examples of formats for the presentations of quality plans.
ISO 10007: 1995	Quality management—Guidelines for configuration management.	Provides guidance on the use of configuration management in industry which is required by the project management system to provide technical and administrative direction and surveillance of the project at different conceptual levels and its interface with other management systems and procedures. It first

(*Contd.*)

Table 13.2 The ISO 9000 family of standards (*Contd.*)

Standard	*Title*	*Contents in brief*
		provides a management overview, then describes the process, organization and detailed procedures. It is applicable to the support of projects from concept through to design, development, procurement, production, installation, operation and maintenance and to the disposal of products. It amplifies the configuration management elements found in ISO 9004/1.
	Quality management and quality system elements	
ISO 9004-1: 1994	Quality management and quality system elements—Part 1: Guidelines for quality management.	Provides guidelines on quality management for use within an organization in the development and implementation of a comprehensive and effective system designed to satisfy the customers' needs and expectations while serving to protect the organization's interest.
ISO 9004-2: 1991	Quality management and quality system elements—Part 2: Guidelines for services.	Provides guidelines for the establishment of a quality system for services with the primary objective of preventing unsatisfactory services.
ISO 9004-3: 1993	Quality management and quality system elements—Part 3: Guidelines for processed materials.	Provides a guide to quality system elements applicable to processed materials, such as bulk products, and discusses means of ensuring effective quality management.
ISO 9004-4: 1993	Quality management and quality system elements—Part 4: Guidelines for quality improvement.	Provides management guidelines for continuous quality improvement within an organization. Describes tools and techniques for a quality improvement methodology based on data collection and analysis.

(*Contd.*)

	Guidelines for auditing quality systems	
ISO 10011–1: 1990	Guidelines for auditing quality system—Part 1: Auditing.	Provides guidelines for establishing, planning, carrying out and documenting audits of quality systems besides providing basic principles, criteria and practices. It allows users to adopt the guidelines to their needs.
ISO 10011–2: 1991	Guidelines for auditing quality system—Part 2: Qualification criteria for quality systems auditors.	Sets out qualification criteria for the selection of auditors to perform quality systems audits in accordance with ISO 10011–1.
ISO 10011–3: 1990	Guidelines for auditing quality system—Part 3: Management of audit programmes.	Provides basic guidelines for managing quality systems audit programmes. These guidelines can be used to establish and maintain an audit programme function when performing quality systems audits in accordance with ISO 10011-1 and ISO 10011–2.
	Quality assurance requirements for measuring instruments	
ISO 10012–1: 1992	Quality assurance requirements for measuring equipment—Part 1: Metrological confirmation system for measuring equipment.	Specifies quality assurance requirements for a supplier to ensure that measurements are made with the intended accuracy, and provides guidelines on implementation; also specifies the main features of the confirmation system for measuring equipment.
	Guidelines for developing quality manuals	
ISO 10013: 1995	Guidelines for developing quality manuals.	Provides guidelines for the development, preparation and control of quality systems procedures, as required by the ISO 9000 family of standards. It can be tailored to the specific needs of suppliers of goods and services.
	Quality vocabulary	
ISO 8402: 1994	Quality management and quality assurance vocabulary.	Defines basic terms under the following headings: general terms; terms related to quality; terms related to quality systems; and terms related to tools and techniques.

Figure 13.1 summarizes the ISO 9000 architecture. As we see, ISO 9001 is the most complex and encompassing of all the three standards of Quality Assurance. It is the most demanding of the three standards to implement, as it requires the greatest commitment of time and resources from an organization. It is also the most expensive and as a result often only appropriate and justifiable for very large organizations to undertake, or companies that must include the design process in their quality management system. ISO 9002 is the most common of the ISO 9000 series and is identical to that of ISO 9001, but does not contain the design and development requirements. This makes the system easier to install but still requires a substantial commitment of time and resources from an organization. The cost of installing this system is less than that of ISO 9001 but still large enough that many organizations may not have the resources to afford such an expense. ISO 9003 is the least complex and easiest to install of the three standards. This system is generally only relevant to simple products and services. It is also an option for those organizations which cannot justify the expense of one of the other systems but still desire a quality management system for their organization. ISO 9004 provides its users with a set of guidelines "by which quality management systems can be developed and implemented".

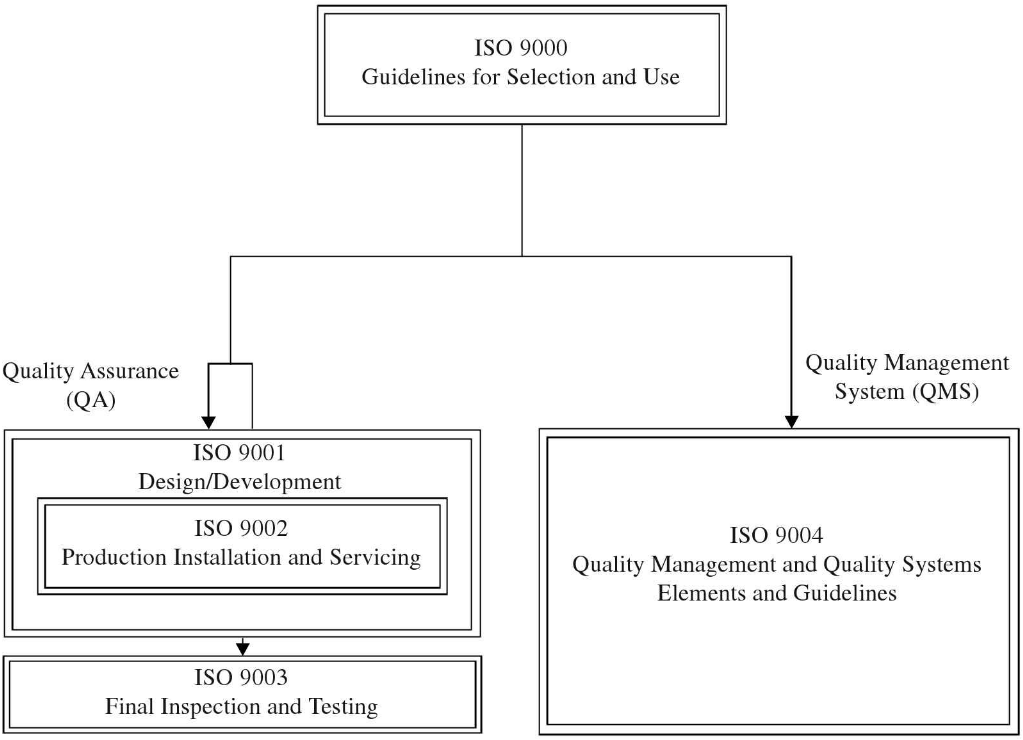

ISO 9000
Guidelines for Selection and Use
Quality Assurance (QA)
Quality Management System (QMS)
ISO 9001
Design/Development
ISO 9002
Production Installation and Servicing
ISO 9003
Final Inspection and Testing
ISO 9004
Quality Management and Quality Systems Elements and Guidelines

Figure 13.1 ISO 9000 architecture.

Note that since ISO 9000 and 9004 are guidelines, it is in fact erroneous to speak of "ISO 9000 Registration/Certification". We can only obtain registration in either 9001, 9002, or 9003.

Further, the series of Standards (9000-9004) are not technical in content in that they do not specify nor set criteria, e.g. minimum purity pH tolerances, hardness requirements, etc. for products. Rather, the quality system standards, 9001, 9002, and 9003 "complement the relevant product or service requirements" given in the technical specifications of the relevant product or service. Each standard focuses on the documentation of operational techniques and managerial activities used to fulfil customer expectations and requirements.

When an organization wants to register for one of the three quality assurance ISO standards (9001, 9002, or 9003), it must first decide (unless otherwise dictated by its customers(s)) which one of the three models is suited best for its needs? As indicated earlier 9003 is a subset of 9002, which in turn is a subset of 9001. The most detailed model for quality assurance, 9001, consists of twenty paragraphs, the 9002 standards consists of eighteen paragraphs whereas 9003 only has twelve. When one recalls that 9003 is the least used, the decision tree consists of a simple binary choice—"Are we involved in Design and Development?" If the answer is "yes" then one should choose 9001. If the answer is "No", then 9002 is very likely to suit the needs. But many a times the decision may not be straightforward.

13.3 QUALITY

ISO 9000 standards, as depicted in Table 14.2, are based on the concept of quality and use of many terms which need some more explanation. We will consider them now.

13.3.1 What is Quality?

Quality has been defined in a number of ways. Some are given below:

- Fitness for use
- Compliance with specific requirements
- Freedom from defects, imperfections, or contamination
- Degree of excellence
- Customer satisfaction
- Delighting customers
- The totality of characteristics of an entity that bear on its ability to satisfy stated and implied needs as defined in ISO 8402 : 1994, Quality Management and Quality Assurance—Vocabulary.

Quality can also be expressed by a simple formula:

$$Q = \frac{P}{E}$$

where P is performance or result and E is customer's satisfaction. When $Q = 1$, the customer's expectations are complied with and full customer's satisfaction is attained.

13.3.2 Quality Policy

This policy explains the overall quality intentions and direction of an organization regarding quality as formally expressed by top management.

13.3.3 Quality Planning

It comprises activities that establish the objectives and requirements for quality and for the application of quality system elements. It covers the following:

- Product planning
- Managerial and operational planning
- Preparation of quality plans
- Making provisions for quality improvement.

13.3.4 Quality Management

ISO 8402:1994 defines quality management as: *All activities of the overall management function that determines the quality policy, objectives and responsibilities, and implements them by means such as quality planning, quality control, quality assurance and quality improvement within the quality system.*

Thus quality management falls within the overall management function of a company. It establishes quality policy, quality objectives, and allocates responsibilities within the organization for achieving the quality policy and objectives.

13.3.5 Quality System

A quality system incorporates the organizational structures, responsibilities, procedures, processes, and resources for implementing quality management. Note that a quality system should only be as comprehensive as needed to meet the quality objectives. For contractual, mandatory and assessment purposes, demonstration of the implementation of the identified elements may be required.

Quality systems organize resources so as to achieve certain objectives by laying down rules and an infrastructure which, if followed and maintained, will yield the desired results. Most elements of the ISO 9001 standard refer to documents that need to be in place, but the extent of the complexity of the documents is left to the professional judgment of the persons creating the system. The most important documents to be developed for a quality system are documents on situations where control is lacking.

13.3.6 Quality Control

ISO 8402:1994 defines quality control as the *operational techniques and activities that are used to fulfil requirements for quality*. This involves both process monitoring and eliminating the causes of unsatisfactory performance at all stages. Thus quality control is a process of maintaining standards and not of creating them. In a simple form, quality control prevents undesired deviations from the planned quality of the product or service being supplied.

13.3.7 Quality Assurance

Quality is not something that can be tacked on to manufacturing or service processes. It is achieved by relying on a philosophy of total integration and control of all the elements of a business. This philosophy is called Quality Assurance. In order to assure quality, it is necessary to ensure that all the requirements for the total presentation are known. In other words, the customers' requirements must be sufficiently detailed to be fully understood by the supplier, so that there are no areas of doubt as to the service requirements.

Quality assurance, therefore, comprises the *planned and systematic activities implemented within the quality system, and demonstrated as needed, to provide adequate confidence that an entity (product or service) will fulfil requirements for quality.*

The ultimate goal of quality assurance for an industry is to develop its image and reputation with regard to:

- Product Quality
- Reliability
- Price
- Delivery
- Customer care

Quality assurance does not just involve regular quality control or inspection checks of the products; it can be an aid to productivity and a means of getting products right—first time, every time.

The background of quality assurance is based on the total integration and control of all elements within a particular area of operation. These elements cover such aspects as administration, finance, sales, marketing, product design and engineering, manufacture, customer aftercare, etc. Quality assurance is, therefore, a management function which cannot be delegated.

Quite often, the means to provide quality assurance have to be built in the process by creating records, documenting plans, documenting specifications and reporting reviews. Such documents and activities also serve to control quality as well as assure it. ISO 9001 provides a model both for obtaining quality for the customer, and for controlling quality for the supplier.

The differences between QC and QA are summarized in Table 13.3.

Table 13.3 Difference between QC and QA

Quality control	*Quality assurance*
Produces results	Is the result of an examination
Detects deviations	Prevents deviations
Can change product quality	Does not change product quality
Concerns the operational means to fulfill quality requirements	Aims at generating confidence that requirements will be fulfilled both within the organization and externally among the customers

13.3.8 The Deming Cycle

W. Edwards Deming, known for introducing statistical quality control to the United States military establishments and to Japanese industries, and author of many books on quality, divides quality control into the following four activities (see Figure 13.2):

- Plan (establish goals, standardize working procedures, and train employees)
- Do (carry out the work according to the plan)
- Check (verify compliance with plans)
- Act (in case of non-compliance, including finding and removing its root cause)

The Deming PDCA (Plan, Do, Check, Act) circle applies to all situations and areas where quality control is wanted. It is a universal model and covers all activities to quality control as well as quality assurance and quality improvement.

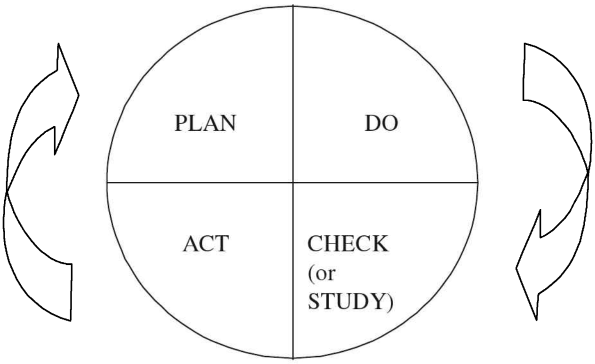


Figure 13.2 The Deming cycle of quality control.

Figure 13.3 shows the Quality Circle indicating steps for each of the four activities of the Deming circle.

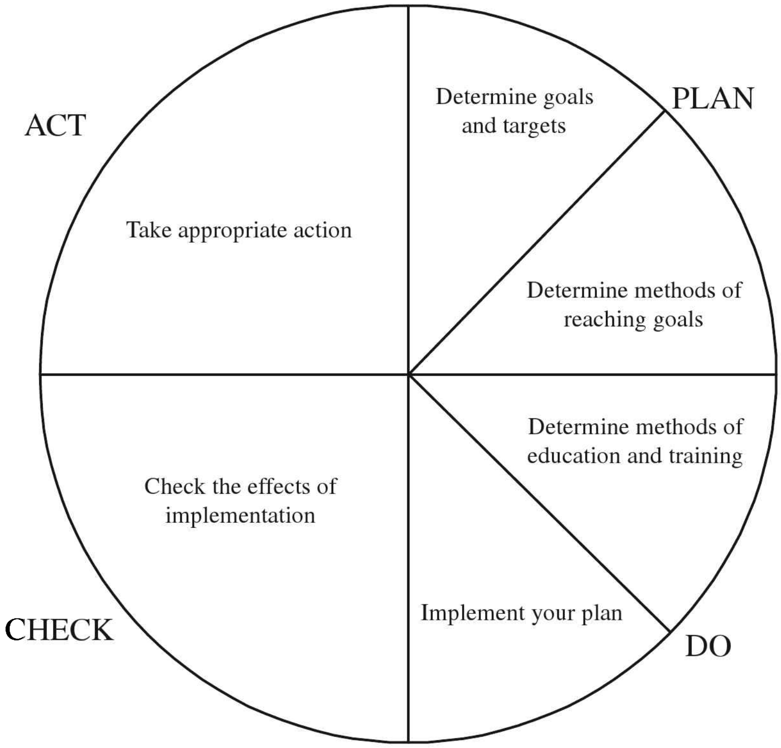

ACT
PLAN
CHECK
DO
Determine goals and targets
Take appropriate action
Determine methods of reaching goals
Determine methods of education and training
Check the effects of implementation
Implement your plan

Figure 13.3 Quality circle.

13.3.9 Quality Loop

This is a conceptual model of interacting activities that influence the quality of a product or service during the various stages ranging from the identification of needs to the assessment of whether these needs have been satisfied. Figure 13.4 illustrates these activities.

13.3.10 Quality Cost

Most organizations do not measure quality costs explicitly, since such costs are not required for the development of financial profit and loss statements, although they can have a significant effect on the profitability of a company. A product does not come into existence as the result of a single activity. There are usually a number of intermediate processes involved before products are dispatched to the customer. While the costs of labour, materials, facilities, machines, etc. are unavoidable, the costs incurred on prevention, detection and removal of errors during production are preventable. Why should the consumer pay for errors? An effective quality management system provides opportunities to keep prevention costs at a reasonable level and helps to avoid unnecessary appraisal and failure costs. It is common to use the PAF (**P**revention, **A**ppraisal and **F**ailure) costing model to evaluate the financial

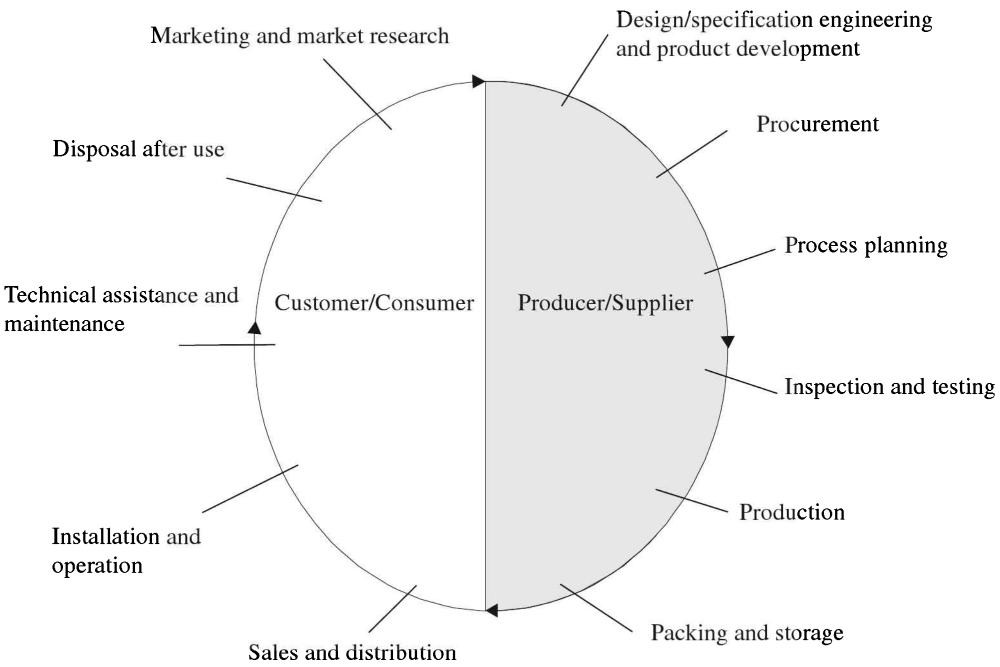


Figure 13.4 Quality loop.

effectiveness of a quality management system. Prevention and appraisal costs are considered investments, while failure costs are regarded as losses.

The component costs include the following:

- ***Prevention cost.*** The cost of efforts to prevent failures. This comprises, for instance, the cost of establishing an effective quality management system including the cost of preparing a quality manual, drawing up procedures, and the like.
- ***Appraisal cost.*** Costs of testing, inspection and examination to assess whether requirements are being fulfilled. This includes the wages of quality control personnel and internal quality auditors, the cost of materials in the testing laboratory, the cost of maintaining and calibrating test equipment, fee paid to certification bodies, etc.
- ***Cost of failure.*** This consists of costs of internal failure and external failure. Internal failure cost results from the failure of a product to meet quality requirements prior to delivery. It covers re-performing a service, reprocessing, rework, retesting, and so on. External failure cost after delivery of a product includes product maintenance, repair, warranties and returns, the cost of recalling the product, liability costs, etc. The lack of goodwill and loss of customers are a very heavy part of this cost.

Quality related costs are generally thought to account for 10–20% of a company's total cost. A typical distribution of this cost is as follows:

- Prevention: 5% of the total quality-related cost
- Appraisal: 30% of the total quality-related cost
- Failure: 5% of the total quality-related cost

Variation of quality cost with defects is shown in Figure 13.5. It is observed that as quality becomes an integral element, the preventive costs stabilize and the total cost declines to the zero defect point. Also the failure costs decline as preventive costs level out.

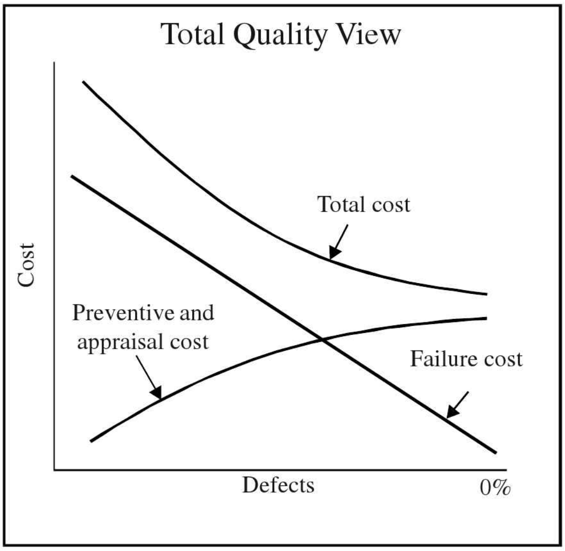


Figure 13.5 Variation of quality cost with defects.

13.3.11 Seven Wastes

Overproduction ("Just-in-case" is Not "Just-in-time".)

Waiting (How long would you wait?)

Transporting (The bigger the gap, the greater the distance.)

Inappropriate processing (Big is not necessarily good.)

Excess inventory (The cost is greater than the money tied-up in it.)

Extra motions (What is needed should be close at hand.)

Defects (Any defect is a waste.)

13.3.12 Quality Improvement

ISO 8402:1994 defines quality improvement as *actions taken throughout the organization to increase the effectiveness of activities and processes in order to provide added benefits to both the organization and its customers*. In simple terms, quality improvement is anything which causes a beneficial change in quality performance.

The transition between where quality improvement starts and quality control begins is where the quality level has been set and mechanisms are in place to keep quality at or above that level. For example, if quality improvement reduces the quality cost from 25% of the turnover to 10% of the turnover, the objective of the quality control is to prevent the quality costs from rising above 10% of the turnover.

13.3.13 Quality Audits

The Quality Audit is defined by ISO 8402:1994 as *a systematic and independent examination to determine whether quality activities and related results comply with planned arrangements and whether these arrangements are implemented effectively and are suitable to achieve objectives.*

Audits and reviews are required to ensure that the quality system is functioning properly. Audits are carried out to ensure that the actual procedures being used do conform to those documented. Reviews, which should be carried out periodically and systematically, reveal whether or not the system is achieving the desired results.

Audits should be carried out by independent investigators since they can be objective. Managers or staff members attempting to audit their own activities are not usually so objective in their appraisal since they are familiar with the *problems* and *traditions* of their activities. Auditors may be *outsiders* brought into the company for the sole purpose of the audit or company employees from other departments within the company.

The audit will include an examination of procedures, records and results, as well as discussions with the staff. The audit process should not be seen confrontational but should attempt to explore the rationale of the operations, the procedures, perceived shortfalls in procedures and attainment and suggestions for possible improvements. The audit should ideally take place away from the work area and in comfortable surroundings. The interviewee should be made to feel important, not under any threat.

The aim of this exercise is for the auditor to learn what is going on so that considered judgments can be made. Only one question should be asked at a time and the auditor should ensure that the answer is complete before moving onto the next question. The auditor should appear to be interested in the answers given. The questions should be phrased in such a way that the answers *yes* and *no* are not possible. Questions should also be clear, i.e. incapable of reinterpretation. Auditors should try to ask questions that make the interviewee to think and explain.

To ensure that a complete appraisal is made, the audit must be thorough, i.e. it should cover all activities within the company's (or department's) normal operating procedures.

A typical audit questionnaire covers the following areas (specific questions will vary according to the type of business the audits are carried at):

- Contract review
- Design control
- Document and data control
- Purchasing
- Control of customer supplied product
- Process control
- Inspection and testing

- Calibration and control of inspection and test equipment
- Inspection and test status
- Control of non-conforming products
- Corrective and preventative action
- Handling, storage, packaging, preservation and delivery
- Control of quality records
- Internal quality audits
- Training
- Servicing
- Statistical techniques

13.3.14 Quality Tools

Following quality tools are very commonly used for the quality management.

- **Statistical Process Control** (SPC). Statistical Process Control is essentially a technique that uses statistical methods for continuous improvement by targeting process variation and thereby reducing costs. One of the main advantages of Statistical Process Control is that information about the process is live, real-time, involves the knowledge of the operators, and any preventive actions can be quickly implemented to prevent waste.
- **Process chart.** It is a list of steps in the manufacture of a product or the delivery of a service. It is useful in identifying the non-business benefit tasks, waste and duplication in the product/service life cycle.
- **Run diagram.** In its simplest form, this is a chart of events plotted against time. It is used to identify if and when problems occur in relationship to time.
- **Histogram.** Similar to the Pareto but shows graphically the relationship between events and frequency of occurrence. When data is classified and plotted it will indicate the most important areas.
- **Pareto analysis.** This is a simple histogram style chart that ranks problems in order of magnitude. It targets the largest, then moves on to the next, then next, and in so doing addresses the area of most benefit first.
- **Ishikawa (fishbone) diagram.** It effectively identifies *cause and effect* in a brainstorm style session that can be easily understood. The main points of influence are the “Ms and Es” (manpower, machines, methods, materials and environment).
- **Measles chart.** Arguably the easiest and most effective tool to chart the frequency of problems and where they occur in or on the physical product. Highlights and targets the areas where the most effective business improvement actions can be placed.
- **Scatter and correlation.** These are used to measure a specific characteristic or an experimental variable against defect levels. Once plotted, the experimental

characteristic can easily be analyzed to show if the desired results were obtained and a relationship identified.

- **Check sheets:** Effectively a predefined audit of the process. Each important area will be assessed for compliance by a predefined check. Often used by pilots prior to flight.
- **Affinity diagram:** It is a brainstorming method used to gather large amounts of data and organize it into natural relationships.

13.4 ELEMENTS OF ISO 9001

Interpretation of quality requirements of document and data control in an electronic environment is given below:

1. **Management responsibility**
   (a) Policy should be clear and defined.
   (b) The responsibility, authority, interrelationships of personnel who manage, perform or verify work effecting quality shall be defined/documented.
   (c) Resource requirements should be identified and assigned to trained personnel.
   (d) Someone should be formally appointed to maintain and report on the quality system.
   (e) Management reviews should be programmed and carried out.
      - Quality related items should be discussed.
      - Actions should be clear, identified and timescaled.
      - Actions should be traceable through meeting-to-meeting until resolved.
2. **Quality system**
   (a) The process should be documented by means of procedure/work instructions as deemed necessary.
   (b) The process should be integrated into the companies hierarchy of quality documentation (ISO 10013 Quality Manuals).
   (c) The effectiveness of the documentation should be monitored.
   (d) That Quality Plan should be available, describing how quality will be assured.
      - Preparation
      - Acquisition of controls, processes, equipment, fixtures, skills
      - Compatibility of design, production process, installation, servicing
      - Updating inspection/testing techniques and new instrumentation
      - Identification of measurement requirements
      - Planning verification stages
      - Defining acceptability criteria
      - Identification of quality records

3. Contract review

(a) The order should be reviewed prior to acceptance to ensure that the:
- Requirements are clear
- Differences between contract and order are resolved
- Supplier has the capability to carry out the work

(b) Any contract amendments should be clearly documented and communicated to all concerned parties

(c) All records of contract reviews should be maintained.

4. Design control

(a) Plans should be available for each design/development stage:
- Defined, described, responsibilities
- Defined activities assigned to qualified personnel
- Reviewed and updated plans as designs evolve

(b) Organizational and technical interfaces for design input groups should be defined.

(c) Design input should include statutory/regulatory requirements and any contract review activities.

(d) Design output should be documented so that it can be verified/validated against:
- Design input requirements
- Reference acceptable criteria
- Crucial design characteristics

(e) The design output documents should be reviewed before release.

(f) At appropriate design stages, formal design reviews should be planned and carried out representation from all concerned areas and specialists as required.

(g) Design verifications should be conducted to ensure that output meets input requirements.

(h) Design validation should be conducted to ensure that product conforms to the defined user needs/requirements.

(i) Design changes should be documented, reviewed and approved by authorized personnel.

5. Document and data control

(a) Documentation/data should be reviewed for adequacy by authorized personnel prior to issue.

(b) A master list/controlling document should be available to identify the current revision of documents/procedures to prevent the inadvertent use of obsolete or superseded documentation.

(c) General controls should ensure that:
  - Pertinent issues of appropriate documentation are available at the place needed
  - Invalid/obsolete documents are removed or managed to prevent unintended use
  - Obsolete documentation retained for *knowledge purposes* are clearly identified

(d) Document/data changes should be reviewed by functions that performed the original approval (unless otherwise documented).

(e) Where practicable, changes to documentation should be identified.

6. Purchasing

(a) Subcontractors should be evaluated to ensure that:
  - They have the ability to meet subcontract requirements
  - Secondary subcontracting is managed/controlled
  - Standard of product is acceptable

(b) Control over subcontractors should be defined.

(c) Acceptable subcontractors should be clearly defined.

(d) Purchasing documents should be clear and include:
  - Type, grade or precise identification
  - Title, applicable specifications, drawings, processes, inspection, data requirements for approval/qualification
  - Number and issue of Quality System applied

(e) Purchasing documents should be reviewed for adequacy.

(f) Purchasing documents should define supplier verification arrangements/visits, etc.

7. Control of customer-supplied product

(a) Documented controls should exist for:
  - Verification
  - Storage
  - Maintenance

(b) Any such product lost, damaged or otherwise unsuitable should be recorded and reported to the customer.

8. Product identification and traceability

(a) The method should be documented for identifying a product by suitable means from receipt to installation/delivery.

(b) The extent of traceability should be defined.

(c) Unique product or batch.

9. Process control

(a) Identification/plan production, installation and servicing processes.

(b) Ensure that processes are carried out under controlled conditions.

- Documented procedures where their absence could adversely effect quality
- Use of suitable equipment and working environment
- Compliance to standards, codes, plans, and procedures
- Monitoring of process parameters and product characteristics
- Approval of processes/equipment
- Defined criteria for workmanship
- Maintenance of equipment to ensure continued process capability
- Process deemed unverifiable until later to be carried out by qualified personnel
- All process, equipment and personnel qualification requirements to be defined

10. Inspection and testing

(a) Inspection and testing activities should be documented.

(b) Inspection and testing requirements, records should be maintained and documented in the Quality Plan or procedures.

- Receiving inspection.
  - Ensure that incoming product is not to be processed/used until confirmed as acceptable (unless under "Positive Recall Controls").
  - Amount/nature of receiving inspection should be defined and the decision traceable.
- In-process inspection and testing
  - Inspection/testing should be carried out to documented procedures.
  - Product should not be released until conformance confirmed (unless under “Positive Recall Procedures”).
- Final inspection and testing
  - Final inspection and testing should be carried out to Quality Plan and procedures.
  - Evidence of finished product conformance to the specified requirements should be identifiable.
  - Ensure that all identified inspection and testing activities are carried out.
  - No completed product should be dispatched until all activities are completed and associated documentation/date authorized.
- Inspection and test records
  - Records should be maintained to provide evidence that the product has been inspected/tested.
  - Records should clearly indicate if product has passed/failed.
  - Records should indicate inspection authority responsible for the release of the product.

11. Control of inspection, measuring and test equipment

(a) The control, calibration, maintenance of equipment should be documented.

(b) Equipment should be used in a manner which ensures that the measurement uncertainty is known and consistent with the required capability.

(c) Test software/comparative references should be checked to ensure capability of verifying acceptability.

(d) Equipment/references should be periodically rechecked for acceptability.

(e) Measurements should be made and accuracy maintained.

(f) All inspection, measuring and test equipment should be clearly identified and maintained under calibration controls.

(g) Calibration status should be identified.

(h) Calibration records should be maintained.

(i) Ensure correct environmental conditions (working and calibration).

(j) Ensure adequate handling, storage, preservation.

(k) Equipment should be safeguarded from unauthorized adjustment .

12. Inspection and test status

- Inspection and test status should be clearly identifiable as defined in the Quality Plan or procedures.

13. Control of non-conforming product

(a) Ensure that non-conforming product is prevented from unintended use.
- Identification
- Evaluation
- Segregation
- Disposition
- Notification

(b) Review and disposition responsibilities should be defined.
- Review in accordance with procedures
- Reworked
- Accepted with concession
- Regraded for alternative use
- Rejected/scrapped
- Description of accepted no-conformance to be recorded
- Repaired/reworked product to be re-inspected

14. Corrective and preventive action

(a) Corrective or preventive action should be taken to eliminate the root cause of potential or actual non-conformances.

(b) Actions taken should be appropriate to the magnitude and risk of the problem encountered.

(c) Any changes to procedures resulting from actions taken should be recorded.

— *Corrective action*

- Effective handling of customer complaints/product non-conformity.
- Investigation of non-conformity and recording of the results.
- Determining corrective action to eliminate root cause.
- Application of controls to ensure that action taken is effective.

— *Preventive action*

- Appropriate sources should be used to eliminate potential causes
- Processes
- Work instructions
- Concessions
- Audit results
- Records
- Service reports
- Customer feedback
- The steps should be determined to deal with any problem requiring preventive action
- Initiation of action and application of controls to ensure effectiveness
- Relevant actions should be fed into the Management Review Structure

15. Handling, storage, packaging, preservation, and delivery

(a) *Handling*

- Methods of handling to prevent damage/deterioration should be defined.

(b) *Storage*

- Designated storage areas should be identified.
- Methods to prevent damage/deterioration should be defined.
- Appropriate methods of authorizing receipt/dispatch should be defined.
- Stock checks should be carried out to determine the condition of stock.

(c) *Packaging*

- Control of packing, packaging, and identification processes should be defined.

(d) *Preservation*

- Appropriate methods for preservation/segregation of product should be defined.

(e) *Delivery*

- Appropriate methods for the protection of the product after final inspection and test should be defined.
- Where contractual, delivery to destination methods should be documented.

16. Control of quality records

(a) Identification, collection, indexing, access, filing, storage, maintenance, and disposition of records should be defined.
(b) All quality records should be legible and stored in such a manner to prevent deterioration.
(c) All quality records should be readily retrievable and activities available to prevent loss.
(d) Retention times for all quality records should be defined.

17. Internal quality audits

(a) Audits should be planned and carried out to determine the effectiveness of the systems.
(b) Audits should be scheduled on the basis of importance/status of the activity to be audited.
(c) Audits should be carried out by qualified personnel independent of the area to be audited.
(d) Results of audits should be recorded and reported to personnel with the authority to take the appropriate corrective action.
(e) Any deficiencies noted during audits should be actioned in a timely manner.
(f) Follow-up activities should be carried out to verify the effectiveness of corrective action taken.

18. Training

(a) Training needs should be identified.
(b) Training plans should be defined.
(c) Specific tasks should be assigned only to appropriate qualified personnel on the basis of:
  - Education
  - Training
  - Experience

(d) Records should be maintained.

19. Servicing

(a) Where contractual performing, verifying and reporting servicing requirements should be defined.

20. Statistical techniques

(a) Any need for statistical techniques for establishing, controlling and verifying process capability/characteristics should be defined.
(b) The application of statistical techniques to be documented and the techniques requirements should be clearly defined.

13.5 TOTAL QUALITY MANAGEMENT (TQM)

The TQM can be considered as a totally integrated effort for gaining competitive advantage by continuously improving every facet of organizational culture. We can say that TOTAL means that every person in the firm is involved (and, where possible, its customers and suppliers), QUALITY means customer requirements are met exactly and MANAGEMENT signifies that senior executives are fully committed

The basic principles of TQM can be classified under ten major headings:

- Leadership
- Commitment
- Total customer satisfaction
- Continuous improvement
- Total involvement
- Training and education
- Ownership
- Reward and recognition
- Error prevention
- Cooperation and teamwork

13.5.1 The Steps to Total Quality Management

Understanding quality

Quality of design may be expressed in terms of how well the product or service is designed to achieve its stated purpose. Design in this context refers to the planning of the product or service. It is not restricted just to the aesthetic considerations. The most important feature of the design relative to achieving the prescribed level of quality is the specification which will clarify the true requirements and capabilities.

Customers expect the product or service to achieve the stated level of *quality of conformance to the design*. Furthermore they would expect a consistent conformance with the agreed specification.

Quality cannot be inspected into the product or service. Inspection/testing/measuring should be carried out to monitor production to see that it does conform to the specification and the fact that testing is an integral part of the quality policy gives confidence to a potential customer. Testing does not actually improve the quality, however, it merely indicates whether or not the required level of quality has been achieved.

Quality can be designed (planned) into the product or service. The designing (planning) of procedures is meant to ensure that a consistent and acceptable level of quality is obtained. This leads to the quality assurance for the ultimate consumer.

Commitment and policy

Commitment to the concept of Total Quality Management must be company- wide. Total Quality Management must be practised constantly. Lip service to the idea will not do.

Quality Policy is a fundamental requirement of any Total Quality Management scheme. The quality policy should be initiated, developed and stated, i.e. it should be made known to all employees. The quality policy may be stated as a slogan, e.g. Quality Right First Time, Rapid Response or Just-in-Time. The quality policy may be amended or updated in the light of experience.

Organization for quality

Quality is the responsibility of everybody—from Chief Executive/Managing Director through production staff to sales and support staff. It is vital to avoid the THEM and US feeling. Every member of the organization must recognize the importance of quality and play his/her part. Necessary action should not be left to THEM (somebody else).

The particular responsibilities of all employees should be defined, as should the lines of responsibility. The exact responsibilities of staff will depend on the nature and size of the company or organization.

Examples of responsibilities are as follows:

Senior executive level. Although commitment to quality is the responsibility of all employees, one senior executive should be designated as directly responsible for quality in the same way that others will be particularly responsible for design, production and sales. If a Director holds the responsibility, the remainder of the workforce will see that quality policy is taken seriously.

First line supervision level. The line manager will be "on the spot" and able to tell whether the supplier/customer interfaces are working correctly. He/she will probably (but not necessarily) be an experienced, long serving company employee with a sound knowledge of the product or service and procedures. The responsibilities of the first line supervisor are to:

- Instruct subordinates in methods and procedures.
- Inform subordinates of likely causes of errors and defects—and prevention methods.
- Supervise/arrange such methods and procedures.
- Initiate any steps thought necessary to improve materials, equipment, methods, and conditions of work.

Employee level. All employees should be sufficiently motivated to:

- Follow agreed written procedures
- Use any materials and/or equipment correctly, as instructed
- Draw attention to any existing or perceived quality problems
- Report all errors, defects, and waste
- Suggest ways in which risk of errors or quality problems could be reduced

- Assist in the training of new employees and young people either as a demonstrator/ trainer or by setting a good example at work

Performance indicators and objectives should be developed if serious quality management is undertaken. Performance indicators are the means by which performance is evaluated. To be meaningful:

- Measurable indicators must lead to quantifiable and tangible objectives. Achievement must be recordable, verifiable and observable, e.g. quantity or quality of output, time schedules, costs, etc.
- Relevant performance indicators must relate to defined areas of responsibility and performance objectives if achievement is to be monitored satisfactorily.
- Relevant performance indicators must be important

Indicators need not be defined for every area of responsibility but should be developed for those activities which have a significant impact on the results.

Action plans should be developed. These include scrutiny of all aspects of:

- Marketing and sales
- Design
- Production operations
- Costs
- Distribution
- After-sales service
- Organizational aspects
- Technology

Action plans should be split into stages:

- Plan
  - — Select a department for study
  - — Identify the best competitor
  - — Identify the benchmarks (criteria)
  - — Decide the information and data collection technology
- Analysis
  - — Compare the company and its competitors using the benchmark data
- Develop
  - — Set performance level objectives
  - — Improve and implement specific action plans
- Review the performance

Once TQM is achieved, the steps to World Class Quality with zero defects can be achieved. Figure 13.6 shows the hierarchy of the Quality system.

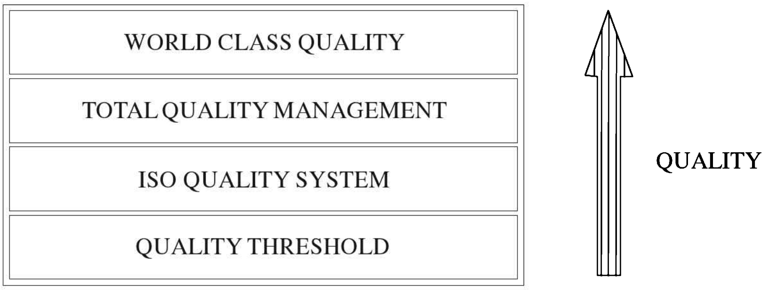


Figure 13.6 Hierarchy of quality system.

13.6 SOME OTHER QUALITY MANAGEMENT STANDARDS

13.6.1 ISO 14000

The ISO 14000 series of environmental management standards was developed by the ISO/TC 207 convened in 1993. Its work was based upon previous consultations that had taken place between the ISO and the IEC's (International Electrochemical Commission) Strategic Advisory Group on the Environment (SAGE) which was itself convened in 1991. It set out to create the framework of the fundamental generic elements that would form the basis for a series of internationally recognized environmental standards, which it completed and published in 1996. The ISO 14000 series of standards represents the essential requirements that every enterprise needs to address in order to control and minimize the impact that its operation, and resulting goods and services, has on the environment. These requirements make up the standards that comprise the environmental management system and their generic nature allow for their application in any type of organization. The management structure that this standard develops within an organization allows the organization to set goals and targets for conforming to the environmental regulations that are required of its industry. Like the ISO 9000 series, this standard has gained acceptance worldwide due to its capacity to be audited by an independent, third party. This allows the required regulatory bodies to objectively measure whether or not the organization is meeting the standards required of them, even though the ISO 14000 series of standards does not specify what those regulations are, since this is normally the job of governments and regulatory bodies. If it were to specify what the regulations were, then a different and unique system would be required for every industry defeating the purpose of having a universally applicable environmental management system. Instead it provides the organization with a management tool that gives it a framework to ensure compliance with all of its external environmental requirements, as well as any internal policies and strategies it has for addressing these issues. It makes sure that these are followed successfully, documented and available for review by interested parties.

13.6.2 ISO 14001

This is a very comprehensive standard and is designed for organizations that have environmental requirements, either internally or externally imposed, with which they need to conform. Often this means providing objective evidence that can be reviewed or audited by appropriate authorities to ensure that the organization is conforming to regulations and that the environmental management system is operating properly. This standard can also be used to demonstrate to customers that all organization's products and services conform to the environmental requirements that customers demand of their suppliers. This can lead to greater opportunity by making one's organization more attractive to customers who have environmental requirements, since having an ISO 14001 system in place often allows one to by-pass the qualification processes. ISO 14001 requires a substantial commitment of time and resources to implement, making it difficult for some organizations to justify its implementation.

13.6.3 ISO 17025

This variation on ISO 9002 was developed specifically for testing and calibration laboratories and up until recently was referred as ISO 9002—Guide 25. It is the entire ISO 9002 standard with additional requirements that are specific to the testing and calibration laboratory industry. These additional requirements are quite rigorous and make implementation of the standard even more challenging than that of the ISO 9002 standard upon which it is based. This standard requires a substantial commitment of time, resources and expense by an organization that undertakes its implementation.

13.6.4 QS 9000

The QS 9000 series of quality management standards was developed by a joint effort of the "Big Three" automakers, General Motors, Chrysler and Ford, and introduced to the industry in 1994. Some heavy truck makers in the United States have also adopted it as well. The program requirements are divided into three sections with the first section being an entire ISO 9001 program with some additional requirements focusing on automotive and truck standards. The second section is titled "Additional Requirements" and contains system requirements that all "Big Three" automakers have adopted and are substantially more rigorous and comprehensive than the ISO 9001 standard. The third and final section is titled "Customer Specific Section" and contains the system requirements that are unique to each and every auto or truck maker that installs this system. It defines the requirements they have developed that are specific to the needs of their respective company. This system is only suited to very large organizations as they are the only ones that can justify QS 9000's enormous cost and requirements of time and resources needed to implement it. While this standard is third-party verifiable, only a very few registrars are even allowed to audit this industry-specific quality management standard.

13.6.5 Japanese Industrial Standards (JIS)

The Japanese Industrial Standards (JIS) were originally developed in 1949 by the Japanese Industrial Standards Committee (JISC) and have been evolving ever since. The JISC has been Japan's answer to ISO in many ways. There has been some harmonization of the JIS with the more internationally recognized ISO 9000 series, i.e. with the JISC incorporating the ISO

9000 series into the JIS in 1991. The JIS are extremely sophisticated and complex and go beyond the requirements of the ISO 9000 series but essentially perform the same quality management function. The JIS are more rigorous and comprehensive in scope making them extremely challenging for an organization to successfully implement. This fact has led to their adoption almost exclusively in Japan and makes their requirement outside of Japan very rare. Organizations that have a JIS certification can be considered to be at least as good if not better than an organization that has an ISO 9000 certification with most JIS systems being closer in scope to ISO 9001.

13.6.6 Q-Base

Q-Base is the highly effective small-to-medium sized business alternative to the ISO 9000 series. Developed by Telarc, the New Zealand's leading quality systems certification body, and completed in 1994, it represents the only internationally recognized quality management system available that is designed specifically for small and medium sized businesses. Q-Base takes the requirements of the ISO 9000 standards (it is derived directly from ISO 9002—the most popular of the ISO 9000 series) and presents them free of jargon and in a manner suitable for use by small enterprises that do not employ quality management professionals. The objective was to develop a program that will identify the basic quality management disciplines that a small company must apply to give its customers confidence in the quality of the goods and services it supplies to them and achieve the cost savings that come from implementation of effective quality systems. Q-Base is increasingly becoming recognized internationally as the standard quality management system for small and medium sized businesses because of its effective combination of value as a system and very low cost of implementation. An operating, certified Q-Base program can be enhanced to full ISO 9002 status using only the company's internal resources, should it ever be required to do so. Every organization can afford not only the cost but also the time and resources required to implement the Q-Base system.

13.6.7 E-Base

E-Base is the highly effective small-to-medium sized business alternative to the ISO 14000 series. Developed by the Canadian management consultant guru David King, and completed in 1999, it represents the only internationally recognized environmental management system available that is designed specifically for small and medium sized businesses. E-Base takes the requirements of the ISO 14000 standards (it is derived directly from ISO 14001) and presents them free of jargon and in a manner suitable for use by small enterprises that do not employ environmental management professionals. The objective was to develop a program that will identify the basic environmental management disciplines that small and medium sized businesses must apply to give their customers confidence that the appropriate environmental standards have been met. Once installed, an E-Base program will deliver a small company environmental management system that is affordable, understandable, installs easily and quickly using only existing corporate resources, and ensures that the company covers all their regulatory, legal, ethical and environmental management requirements. The

E-Base program is designed to be installed as a "stand-alone" system, or it may be integrated into an existing quality management program. In addition, an operating, certified E-Base program can be enhanced to full ISO 14001 status using only the company's internal resources, should it ever be required to do so.

13.7 ISO 9000:2000 UPDATE

ISO directives require the standards to be revised on a regular basis. Throughout the revision process, worldwide users of the standard provide feedback on the requirements and the format and text of the documents. In 1994, the ISO 9000 standard was changed, but only around the edges. Now for 2000, ISO 9000 has again been revised. But the new standard, called ISO 9000:2000, is a radical revision.

The key standards within the ISO 9000 family of standards (ISO 9001:1994, ISO 9002: 1994 and ISO 9003:1994) have been merged into a single ISO standard, ISO 9001:2000. The new series is as follows:

- ISO 9000:2000 (QMS—Fundamentals and vocabulary)
- ISO 9001:2000(QMS—Requirements)
- ISO 9004:2000(QMS—Guidance for performance improvement)

ISO 9000:2000 has a new process oriented structure. It includes a process model based on the Plan-Do-Check-Act cycle which outlines the product and/or service cycle and the management control cycle. Figure 13.7 shows the process model for ISO 9000:2000.

In comparison to its predecessor, ISO 9000:2000 actually features several quantum leaps forward. Following are the key changes.

Voice of the customer

The graphic depiction of the process model in Figure 13.7 shows the power of the customer in the new standard. It is bookended by two drivers: "Customer Requirements" drive the input and "Customer Satisfaction" drives the output. The organization needs methods in place to describe and monitor the needs and desires of each customer for each order, and needs processes and procedures in place to measure and analyze customer satisfaction.

Continual improvement

Under ISO 9000:2000, it is not enough for an organization simply to measure customer satisfaction; it needs to improve the level of satisfaction. It also has to measure and improve internal processes. Continual, or continuous, improvement is a core theme in the new version of ISO 9000 and inherent in the model's PDCA structure.

Continual improvement is one of the essential goals of quality. Continually improving defect rates, with consequent increased customer satisfaction, is a core value of Deming and TQM. Although many believe that continual improvement was always implied in ISO 9000, continual improvement is now a clearly defined requirement throughout the 2000 revision.

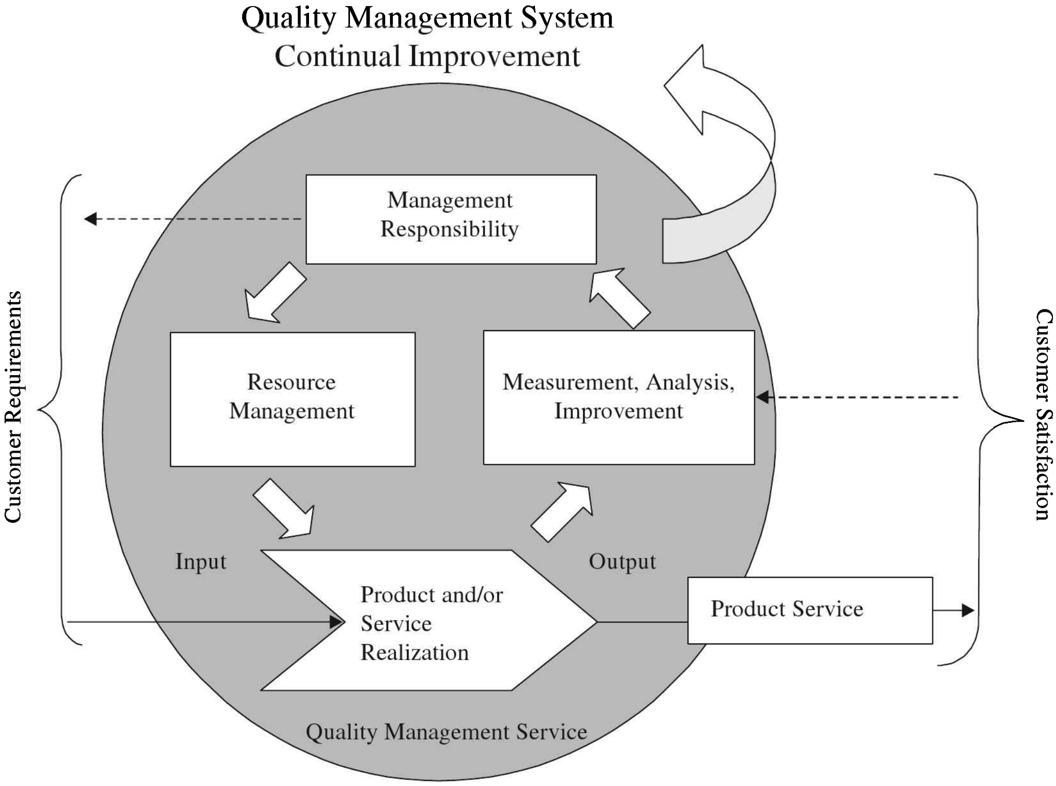

Quality Management System
Continual Improvement
Customer Requirements
Management Responsibility
Resource Management
Measurement, Analysis, Improvement
Input
Product and/or Service Realization
Output
Product Service
Customer Satisfaction
Quality Management Service

Figure 13.7 Process model for ISO 9000:2000.

Management responsibility

Management's role in the previous ISO 9000 was that it was required to establish quality policy, commit adequate resources, conduct a management review and appoint a management representative to supervise the QMS. But, for the most part, under ISO 9000:1994, the QMS was largely the responsibility of quality professionals. Executive management plays a far more central role with the new standard. In ISO 9000:2000, management responsibility is expanded so that management presides over a multi-step version of the Plan-Do-Check-Act process. This process includes the following requirements.

Step 1. Policy. Management is obligated to establish an appropriate quality policy that incorporates a commitment to continual improvement and meeting customer requirements.

Step 2. Objectives. This policy must establish a framework for reviewing quality objectives that are set "at relevant functions and levels within the organization."

Step 3. Planning. Objectives are set in conjunction with a plan that identifies the activities and resources necessary to achieve it. The planning "shall be consistent with other requirements of the quality system."

Step 4. Quality-management system. Management is responsible for the organization establishing a QMS as a means of implementing the quality policy with its associated objectives and plans, as well as the requirements of the standard.

Step 5. Management review. Although ISO 9000:1994 required a management review, this element is much expanded in the new version. In this new model, management's check of the QMS specifically includes a review of policy and objectives to find opportunities for continual improvement. On the basis of this review, management is required to take actions—among other things—"relating to the improvement of the quality-management system."

This process, which begins with setting policy and moves to management review with continual improvement as the output, is similar to the way the PDCA process is outlined in ISO 14001, the Environmental Management System Standard. The compatibility between ISO 14001 and the new ISO 9000 should permit organizations to develop complementary systems for the two standards. (Appendix A of ISO 9001:2000 offers tables that show the correspondence between ISO 9001 and ISO 14001.)

Resource management

The 1994 standard contains a paragraph (4.1.2.2) that requires management to provide necessary resources. It also contains requirements for training (4.18). These kernels are expanded in the 2000 revision to be one of the four clauses, "6.0 Resource Management," in the Process Model. This section spells out a range of specific resources, including human resources that management leadership must provide or make available, i.e. adequate numbers of competent people, the training necessary to ensure competence, infrastructure, work environment, suppliers and partners, and financial resources.

13.7.1 Quality Management Principles

The advances in ISO 9000: 2000 are embodied in the standard's following eight Quality Management Principles. They emphasize the importance of customer focus, leadership, involvement of people at all levels, the process approach, systems and objectives, continual improvement, importance of accurate data and analysis, and the significance of supplier relationships—a mirror image of the focus on the customer.

13.7.2 Breakdown of Sections in ISO 9000:2000

In place of the 20 elements in ISO 9000:1994, there are now 8 sections in ISO 9000:2000. The breakdown is as follows:

Section 1: Scope

1.1: General
    (a) demonstrate ability
    (b) address customer satisfaction

1.2: Permissible exclusions:
- (a) requirements that do not affect the organization's ability to provide acceptable product
- (b) requirements that do not affect the organization's responsibility to provide acceptable product

Section 2: Normative Reference (very similar to ISO 9001:1994)

Section 3: Terms (refers us to ISO 9000:2000)

Section 4: Quality Management

4.1: General Requirements
- (a) identify processes
- (b) determine sequence and interaction
- (c) determine criteria and methods
- (d) ensure availability of information
- (e) measure, monitor and analyze and implement action necessary to achieve results

4.2: General Documentation Requirements
- (a) necessary procedures
- (b) documents required to ensure effective operation and control of processes

The extent of the documentation will be dependent on:
- (a) size and type of organization
- (b) complexity and interrelation of processes
- (c) competence of personnel

Section 5: Management Responsibility

5.1: Management Commitment
- (a) communicating to organization importance of customer, regulatory and legal requirements
- (b) establishing the quality policy and objectives
- (c) conducting management reviews
- (d) ensuring necessary resources

5.2: Customer Focus

Ensure that customer needs and expectations are determined, converted into requirements and fulfilled with the aim of achieving customer requirements

5.3: Quality Policy
- (a) appropriate to the organization’s purposes
- (b) including commitment to requirements and continual improvement
- (c) framework for establishing and reviewing quality objectives
- (d) communicated and understood at appropriate levels of the organization
- (e) reviewed for continuing suitability

5.4: Planning
  5.4.1 Quality objectives
  5.4.2 Quality planning

5.5: Administration
  5.5.1 General
  5.5.2 Responsibility and authority
  5.5.3 Management representative
  5.5.4 Internal communication
  5.5.5 Quality manual
  5.5.6 Control of documents
  5.5.7 Control of quality records

5.6: Management Review
  5.6.1 General
  5.6.2 Review input
  5.6.3 Review output

Section 6: Resources

6.1: Provision of Resources

6.2: Human Resources
  6.2.1 Assignment of personnel
  6.2.2 Training, awareness and competency

6.3: Facilities

6.4: Work Environment

Section 7: Product Realization

7.1: Planning of Realization Process
  (a) quality objectives for the product, project or contract
  (b) the need to establish processes and documentation, and provide resources and facilities specific to the product
  (c) verification and validation activities, and the criteria for acceptability
  (d) the records that are necessary to provide confidence of conformity of the processes and resulting product.

7.2: Customer-Related Process
  7.2.1 Identification of customer requirements
  7.2.2 Review of product requirements
  7.2.3 Customer communications

7.3: Design and/or Development
  7.3.1 Design and/or development planning
  7.3.2 Design and/or development inputs

- 7.3.3 Design and/or development outputs
- 7.3.4 Design and/or development review
- 7.3.5 Design and/or development verification
- 7.3.6 Design and/or development validation
- 7.3.7 Control of design and/or development changes

7.4: Purchasing

- 7.4.1 Purchasing control
- 7.4.2 Purchasing information
- 7.4.3 Verification of purchased product

7.5: Production & Service Operations

- 7.5.1 Operations control
- 7.5.2 Identification and traceability
- 7.5.3 Customer property
- 7.5.4 Preservation of product
- 7.5.5 Validation of processes

7.6: Control of Measuring and Monitoring Devices

- (a) Should be calibrated and adjusted periodically or prior to use, against devices traceable to international or national standards; where no such standards exist, the basis used for calibration shall be recorded.
- (b) Should be safeguarded from adjustments that would invalidate the calibration;
- (c) Should be protected from damage and deterioration during handling, maintenance and storage.
- (d) Should have the results of their calibration recorded
- (e) Should have the validity of previous results re-assessed if they are subsequently found to be out of calibration, and corrective action taken.

Section 8: Measurement, Analysis, Improvement

8.1: Planning

8.2: Measurement and Monitoring

- 8.2.1 Customer satisfaction
- 8.2.2 Internal audit
- 8.2.3 Measurement and monitoring of processes
- 8.2.4 Measurement and monitoring of product

8.3: Control of Nonconformity

8.4: Analysis of Data

- (a) Customer satisfaction and/or dissatisfaction
- (b) Conformance to customer requirements
- (c) Characteristics of processes, product and their trends
- (d) Suppliers

8.5: Improvement
- 8.5.1 Planning for continual improvement
- 8.5.2 Corrective action
- 8.5.3 Preventive action

Benefits of the revised standards

There are a number of major benefits with the revised standards:

- More applicability to all product and/or service sectors and to all sizes of companies
- Simpler to use, clearer in language, more readily translatable, and more easily understandable
- Clearer connection of quality management systems to company processes
- Provision of a natural evolutionary approach to improved company performance
- More orientation towards continual improvement, customer satisfaction and resource management
- Compatibility with other management standards such as ISO 14000
- Provision of a consistent basis to address the needs and interests of companies in specific sectors (e.g. medical devices, telecommunications, automotive, etc.)
- The concept of the consistent pair—with ISO 9001 covering the minimum requirements and ISO 9004 providing guidance for going beyond these requirements to further improve company performance and customer satisfaction.

QUESTIONS

13.1 What is the purpose of ISO 9000 Quality Management System? In spite of product standards prevalent in country, what was the necessity of introducing ISO 9000. What are the Indian equivalents of ISO 9000:1994?

13.2 Draw the ISO 9000 architecture. Why can an organization obtain certification in either ISO 9001, ISO 9002 or ISO 9003 and not in ISO 9000 and ISO 9004?

13.3 What is Quality? Explain Quality Policy, Quality Planning, Quality Management and Quality System.

13.4 What is meant by Quality Control and Quality Assurance? Differentiate between the two.

13.5 Explain the Deming cycle.

13.6 Draw the Quality Loop and label all the segments.

13.7 Which model is commonly used to evaluate financial effectiveness of a Quality Management System? Draw the *cost* versus *defects* curves to show the variation of quality costs with defects.

13.8 (a) What are the seven wastes in the domain of quality?

(b) What is meant by Quality Audit? What are the typical areas covered by Quality Audit Questionnaire?

13.9 What are the commonly used Quality Tools used in a Quality Management System?

13.10 Write all the 20 elements of ISO 9001 Quality Management System in sequential order.

13.11 If your organization/institute is to go for ISO 9001 certification, what are the various elements required for implementing the system? Explain any five elements.

13.12 What should be the mission statement of your organization/institute?

13.13 Sketch the diagram of an organization from the stage of being uncommitted to the status of a world-class company.

13.14 What is meant by Total Quality Management (TQM)? What are its basic principles? Explain the underlying steps to TQM.

13.15 In ISO 9000:2000 standards, what are the changes in basic definitions of ISO 9000:1994?

13.16 List the four major and key changes in the ISO 9001:2000 standard from the 1994 standard.

13.17 Draw the Process Model for ISO 9000: 2000 from the standpoint of customer focus and label all the parts of the model.

13.18 What are the management responsibilities in ISO 9000:2000? List them.

13.19 What are the benefits of ISO 9001:2000 standard over the 1994 standard?

SECTION III

Instrumentation Technology

Chapter 14

Industrial Communication Techniques

14.1 INTRODUCTION

Communications plays a very important role in industry. There are many examples:

(a) Transfer of temperature information from the sensor to the input unit, and from the input unit to the programme calculating the control outputs.
(b) Transfer of instructions from the programme to the output unit, and from the output unit to a frequency converter.
(c) From a programmable logic controller to the robot controller to convey the information that the part has arrived at the work cell.
(d) From the controller in the air-conditioning system to the motorventile in the ventilation tube.
(e) In the plant control room, process information is made available on the operator's monitor for operation instructions or such other special instructions from the intranet.

For the communications there are a lot of different kinds of solutions. The simplest one is a connection, where an input board in the robot controller reads the one bit ON/OFF information from the relay output of a programmable logic controller. In the most advanced solution, the control systems share information in a factory-wide network. The network consists of many hierarchy levels (Figure 14.1).

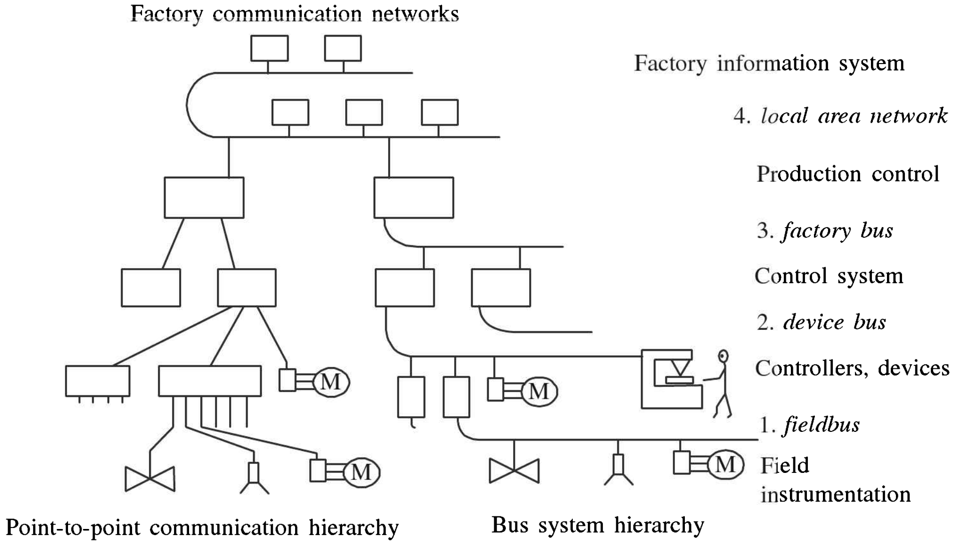


Figure 14.1 Hierarchy of factory communication.

The selection of a communication system is based on:

- the maximum response time required for the information
- the importance of the information for the safe operation of the equipment
- the amount of information to be transferred
- maximum distance
- the properties of the devices to be connected.

For example, the information about the trolley arriving at the end of its track cannot be transmitted through a local area network. The information is only a few bits. The transmission time in a local area network is anyway too long because making the connection takes too long time. The trolley will hit the end of the track before the information will arrive at the controller. The local area network can be used for transferring files containing statistics about the operation hours of a production line. The time required for opening the connection has no role in transferring this file.

The evolution in communication is the result of a technological evolution, which can be traced back to the beginning of the century, to the industrial revolution, and which is still going on. However, in the field of communication, we witness a path that goes in the reverse direction compared to the evolution of, say, mechanical and electronic engineering: from the most abstract layers of data transfer first, then to the application of control and automata theory, and finally to that of sensors and actuators (see Figure 14.2).

With the introduction of steam power, a whole series of clever inventions appeared at the beginning of the century, together with ingenuous systems with levers or other actuators to control and automate processes. Such mechanisms have evolved into cam-driven systems, still used today. However, a brand new technology, based on pneumatics and vacuum, became rapidly popular: thanks to pistons, more flexibility is gained. This was the first step to a

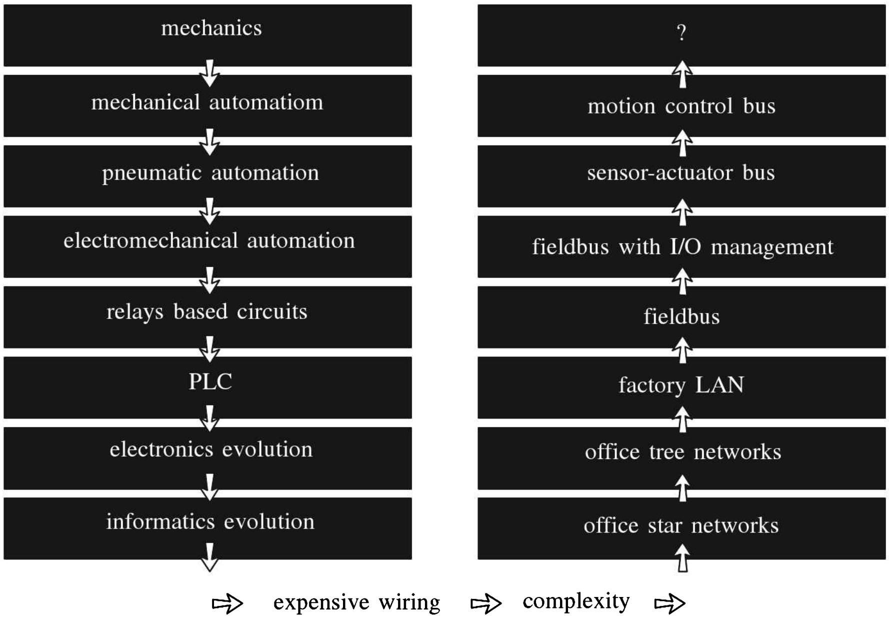


Figure 14.2 Evolution of technology in control systems.

technology that processes input parameters to transform them in logical control signals, which led to the idea of automating processes by combining signals through pneumatic valves.

Electricity has brought all that was needed to reduce volume and complexity of pneumatic systems, without any sacrifice in flexibility. Electric relays have let a great number of pneumatic valves to be substituted by circuits.

A leap in flexibility was made possible by the construction of *programmable automata*, realized with electronics. Such machines, from the first automata which simply replaced existing relays, have grown to become real computers processing an increasing number of data. In general, the introduction of computer hardware and software has contributed, on the one hand, more system complexity, with more demanding data processing, requiring increasingly large and expensive cables and connections; but on the other, computer technology has allowed the development of communication systems.

A first step has consisted in letting computers communicate with each other, thanks to a central node, or *star-like net*, with a point-to-point connection based on RS-232. One of the best known examples is that of time-sharing systems with VAX machines connected to VT-100 terminals—a basic standard for terminals. However, with the technological breakthroughs, the star-shaped networks have disappeared in favour of *tree-structured* nets—the main difference being that users are connected to the same cable and the transferred data are time-multiplexed.

With the improved communication protocols, it is now possible to reduce the time needed for the transfer (and simultaneously ensure better data protection), guarantee the synchronization of processes coming from different production cells, and satisfy the requirements of critical signals in the field bus. Though low level components for field buses, that of sensors/actuators, have been in existence for a few years, they have become popular only in recent times. Buses for the most critical tasks, such as synchronization of movements, are just about to emerge in industrial setups.

The latest communication technique is the digital communication. Industrial networks that transmit data using digital signals often are an integral part of a data acquisition or process control solution. A basic understanding of the network technologies that are available for various applications is required to make the best implementation decisions—decisions that can have a profound effect on the ability to adapt to ever-changing technologies.

For example, the type of network(s) or network products selected for a data acquisition application can greatly affect cost/benefit evaluations for future projects. Until recently, fibre optics was an expensive option that seemed like overkill for most applications. But with the information load now likely to flow between nodes on an enterprise network, fibre looks attractive.

Network technologies come in a bewildering array of options, and certain segments of this technology are changing at an incredible rate. The average user might spend weeks researching the various ways to build or improve a network, just to find that once a decision is made and products are purchased, the next wave of bigger, better, faster technology is now available.

A well-designed, integrated solution to data transmission will yield a competitive advantage to any industrial enterprise. Users in all aspects of a business should be able to obtain plant and business data from any physical node, local or remote. *Gluing* pieces of new and/or existing networks together is becoming more feasible with the use of bridging, routing, and media conversion technologies that link local area, wide area, and industrial networks together. And with Internet and wireless technologies, data transmission over large geographic areas is increasingly feasible.

14.2 THE OSI NETWORK MODEL

The OSI (Open System Interconnection) model of ISO (International Standard Organization) is the official reference for communication and is a reference basis (i.e. a model) to identify and establish a classification of the different functions of communication systems. This model represents therefore only a reference point, and not a set of prescriptions for a system to adhere to. This model defines seven layers of functions (Figure 14.3) in a communication system. It is not compulsory to have all of them in a given system, as depending on the application, some middle layers may be missing.

We will briefly illustrate the role that each layer plays in the communication.

- **The Application layer** (Layer 7) is the most intuitive because it is what the user sees. It represents the problem the user wants the system to solve. Internet browsers and e-mail programs are good examples. They allow the user to input and read data while connected between a client PC and a server somewhere on the Internet. In an industrial application, a program on a programmable logic controller (PLC) might control a smart valve.
- **The Presentation layer** (Layer 6) performs formatting on the data going to and from the application. This layer performs such services such as encryption, compression,

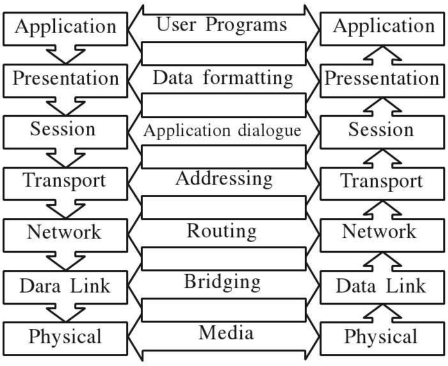


Figure 14.3 The seven-layer OSI network model.

and conversion of data from one form to another. For example, an application (one layer up) might send a time stamp formatted in 12-hour time: 01:30:48 p.m. A more universal representation is 13:30:48, in 24-hour time, which can be accepted or presented at the next node's application in the form that it needs. One advantage to presentation services is that they help eliminate overhead, or embedded services, in application programs.

- **The Session layer** (Layer 5) establishes the connection between applications. It also enforces dialogue rules, which specify the order and speed of data transfer between a sender and a receiver. For example, the session layer would control the flow of data between an application and a printer with a fixed buffer, to avoid buffer overflows. In the time stamp example, once the data is presented in 24-hour time format, an identifier and length indicator are appended to the data string.
- **The Transport layer** (Layer 4) is essentially an interface between the processor and the outside world. It generates addresses for session entities, and ensures that all blocks or packets of data have been sent or received. In the time stamp example, an address for each session entity (sender and receiver) and checksum are appended to the block generated by the session layer.
- **The Network layer** (Layer 3) performs accounting, addressing, and routing functions on messages received from the transport layer. If the message is lengthy, this layer will break it up and sequence it over the network. This layer also uses a network routing table to find the next node on the way to the destination address. In the time stamp example, a node address and sequence number are appended to the message received from the session layer.
- **The Data Link layer** (Layer 2) establishes and controls the physical path of communication from one node to the next, with error detection. This layer performs media access control (MAC) to decide which node can use the media and when. The rules used to perform these functions are also known as protocols. Ethernet and token ring contention are examples of protocols. In the time stamp example, a header and trailer are appended to the message received from the transport layer for flagging the beginning and end of the frame, type of frame (control or data), checksum, and other functions.

- **The Physical layer** (Layer 1) is perhaps the most conspicuous layer from a cost point of view. It is relatively easy to understand the labour and material costs of pulling cables, along with a physical infrastructure (conduits, raceways, and duct banks) for maintaining cable integrity. This layer does not add anything to the message frame. It simply converts the digital message received from the data link layer into a string of ones and zeroes represented by a signal on the media. One example is RS-232.

14.3 PHYSICAL LAYER OPTIONS

There are a number of implementations of the physical layer. Network devices allow a wide range of connectivity options. Some networks are well defined using the OSI model, where cables, bridges, routers, servers, modems, and PCs are easily identified. Sometimes only a few devices are linked together in some kind of proprietary network, or where the network services are bundled in a black-box fashion with the device.

14.3.1 Network Topologies

The term topology refers to the method used to connect components on a network. The most common topologies are *ring*, *bus*, and *star* topologies (Figures 14.4, 14.5, and 14.6), but they can take on the appearances of each other and still maintain their characteristics. For example,

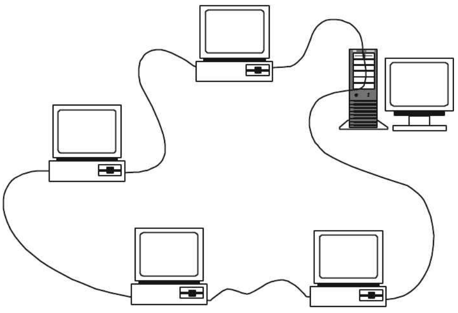


Figure 14.4 Ring topology.

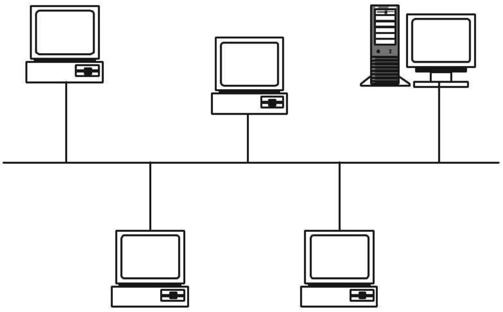


Figure 14.5 Bus topology.

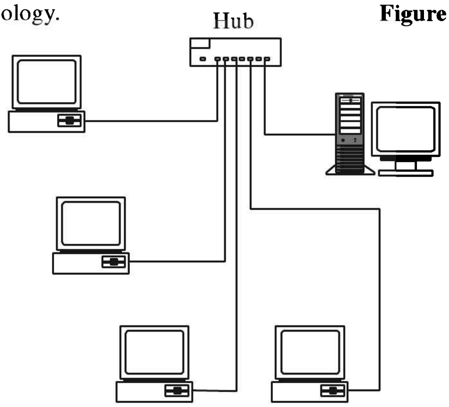


Figure 14.6 Star topology.

a token ring network segment can be wired in a star configuration, where the components are cabled back to a hub where the ring is *inside* the hub. This allows a common wiring closet for a given building or area with home-run cable pulls for each component. A hub offers the advantages of centralized maintenance and configuration control.

14.3.2 Interface Standards

The Electronics Industry Association (EIA) has produced standards for RS 485, RS 422, RS 232 and RS 423 that deal with data communications. The Electronics Industry Association is an industry group based in Washington, DC, USA, which recommends and publicizes standards related to electronic components and communications. Standard number 232, officially called the "Interface between Data Terminal Equipment and Data Communications Equipment Employing Serial Binary Data Interchange", specifies RS-232. The prefix "RS" indicates Recommended Standard. The third revision is denoted by adding C as the last character to RS 232 (so it is represented as RS 232C).

These are the most common serial data exchange interfaces used for connecting two or more devices together. All three interfaces use data terminal equipment (DTE) and data communication equipment (DCE) terminology (Figure 14.7). The DTE is the component that wants to communicate with another component somewhere else, such as a PC communicating with another PC. The DCE is the component actually doing the communication, or, performing the functions of the generator and receiver discussed in the standards. A modem is a common example of a DCE.

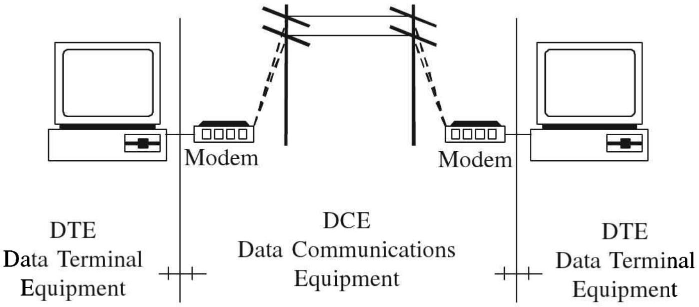


Figure 14.7 Serial data transmission.

The interfaces between DTE and DCE can be categorized by mechanical, electrical, functional, and procedural aspects. Mechanical specifications define types of connectors and numbers of pins. Electrical specifications define line voltages and waveforms, as well as failure modes and effects. Functional specifications include timing, data, control and signal grounds, and which pin(s) the functions are to use. The procedural interface specifies how signals are exchanged.

14.3.3 RS 232

Electronic data communications between elements generally fall into two broad categories—

single-ended and differential. RS 232 provides a single-ended transmission and was introduced in 1962. The specification allows for asynchronous serial data transmission from one transmitter to one receiver at relatively slow data rates (up to 20K bits/second) and short distances (up to 50-ft @ the maximum data rate). The word serial means, that the information is sent one bit at a time. Asynchronous tells us that the information is not sent in predefined time slots. Data transfer can start at any given time and it is the task of the receiver to detect when a message starts and ends. Asynchronous communication has some advantages and disadvantages.

With synchronous communication, a clock or trigger signal must be present which indicates the beginning of each transfer. The absence of a clock signal makes an asynchronous communication channel cheaper to operate. Less lines are necessary in the cable. A disadvantage is that the receiver can start at the wrong moment receiving the information. Resynchronization is then needed which costs time. All data received in the resynchronization period is lost. Another disadvantage is that extra bits are needed in the data stream to indicate the start and end of useful information. These extra bits take up bandwidth.

Character frame

Serial communication using RS 232 requires four parameters: the *baud rate* of the transmission, the number of *data bits* encoding a character, the sense of the optional *parity bit*, and the number of *stop bits*. Each transmitted character is packaged in a character frame that consists of a single start bit followed by the data bits, the optional parity bit, and the stop bit or bits. A typical character frame encoding the letter "m" is shown in Figure 14.8.

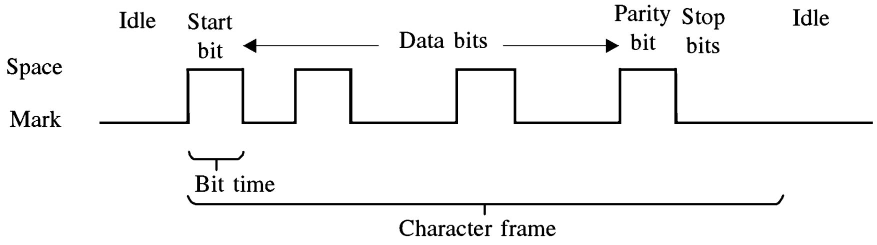


Figure 14.8 A typical character frame.

RS 232 uses only two voltage states, called MARK and SPACE. MARK is a negative voltage and SPACE is positive. The allowed voltage range for SPACE is > +3 V (Logic 0) and for MARK is < –3 V (Logic 1). The output signal level usually swings between +12 V and –12 V. The *dead area* between +3 V and –3 V is designed to absorb line noise.

Data bits are sent with a predefined frequency, the baud rate. Both the transmitter and receiver must be programmed to use the same bit frequency. After the first bit is received, the receiver calculates the moments at which the other data bits will be received. It will check the line voltage levels at those moments.

Start bit. A start bit signals the beginning of each character frame. It is a transition from negative (MARK) to positive (SPACE) voltage; its duration in seconds is the reciprocal of the baud rate. If we are transmitting at 9600 baud, then the duration of the start bit and each subsequent bit will be about 0.104 ms. The entire character frame of eleven bits would be transmitted in about 1.146 ms.

Data bits. Data bits are transmitted "upside down and backwards." That is, inverted logic is used and the order of transmission is from least significant bit (LSB) to most significant bit (MSB). To interpret the data bits in a character frame, we must read from right to left, and read 1 for negative voltage and 0 for positive voltage. For the frame of Figure 14.8, this yields 1101101 (binary) or 6D (hex) which corresponds to the letter "m".

Parity bit. An optional parity bit follows the data bits in the character frame. The parity bit, if present, also follows inverted logic: read 1 for negative voltage and 0 for positive voltage. The parity bit is included as a simple means of error checking and is set depending upon whether the parity of the transmission is even or odd. Suppose the parity is chosen to be odd. The transmitter will then set the parity bit in such a way as to make an odd number of 1s among the data bits and the parity bit. The transmission in the figure above uses odd parity. There are five 1s among the data bits, already an odd number, so the parity bit is set to 0. The transmitter calculates the value of the bit depending on the information sent. The receiver performs the same calculation and checks if the actual parity bit value corresponds to the calculated value.

Stop bits. Suppose that the receiver has missed the start bit because of noise on the transmission line. It started on the first following data bit with a space value. This causes garbled data to reach the receiver. A mechanism must be present to resynchronize the communication. To do this, framing is introduced. Framing means, that all the data bits and parity bit are contained in a frame of start and stop bits. The period of time lying between the start and stop bits is a constant defined by the baud rate and number of data and parity bits. The start bit has always space value, and the stop bit always marking value. If the receiver detects a value other than marking when the stop bit should be present on the line, it knows that there is a synchronization failure. This causes a framing error condition. The receiving device then tries to resynchronize on new incoming bits. For resynchronizing, the receiver scans the incoming data for valid start and stop bit pairs. This works, as long as there is enough variation in the bit patterns of the data words. For example, if data value zero is sent repeatedly, resynchronization is not possible. The stop bit identifying the end of a data frame can have different lengths. Actually, it is not a real bit but a minimum period of time the line must be idle (marking state) at the end of each word. This period can have three lengths: the time equal to 1, 1.5 or 2 bits. 1.5 bits is only used with data words of 5 bits length and 2 only for longer words. A stop bit length of 1 bit is possible for all data word sizes.

Connectors and signal lines

DB-25 connector (Male/Female) is the standard RS 232 connector, with enough pins to cover all the signals specified in the standard. Figure 14.9 shows the images of this connector along with the Hood (Cover).

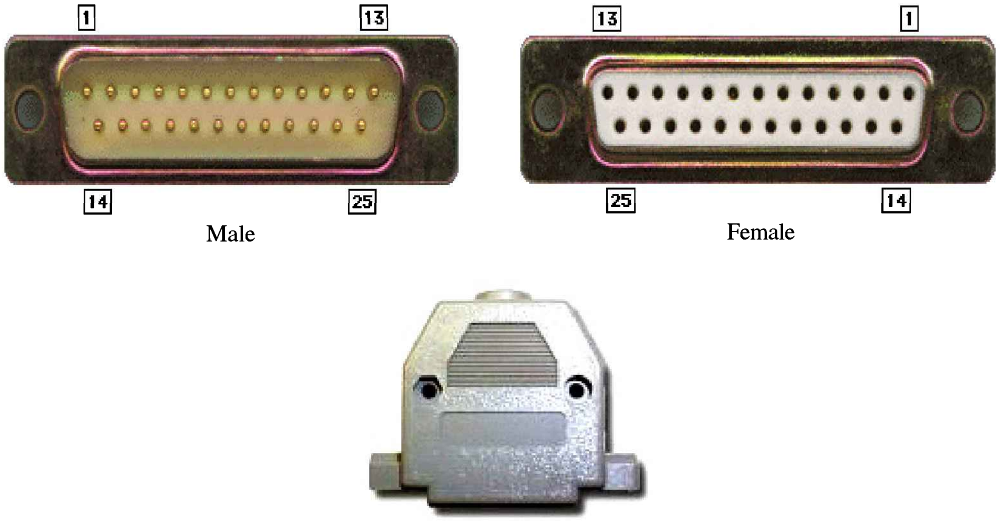


Figure 14.9 DB-25 connector of RS 232.

Another connector DB-9 is occasionally found on smaller RS 232 lab equipment. It is compact, yet has enough pins for the *core* set of serial pins (with one pin extra). The DB-9 pin numbers for transmit and receive (3 and 2) are opposite of those of the DB-25 connector (2 and 3). Figure 14.10 shows the images of the male and female versions of the DB-9 connector with pin numbers.

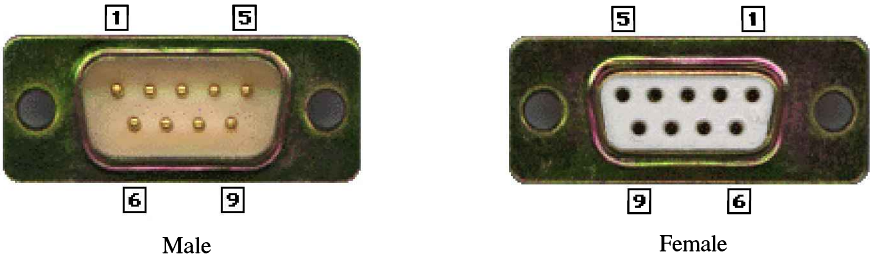


Figure 14.10 DB-9 connector of RS 232.

Figure 14.11 shows the pin assignments of DB-25 and DB-9 connectors.

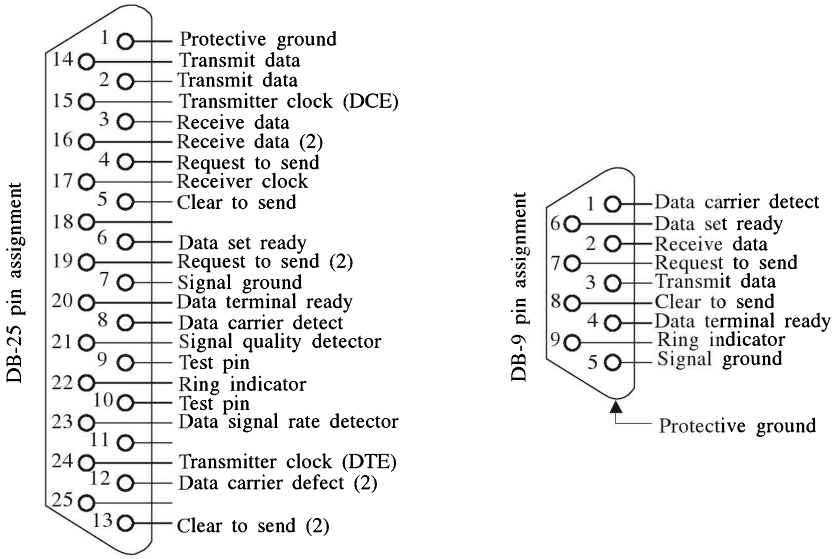


Figure 14.11 DB-25 and DB-9 pin assignments.

Table 14.1 lists the *core* set of pins that are used for most RS 232 interfaces.

Table 14.1 The *core* set of pins used for most RS-232 interfaces

DB-25	*DB-9*	*Signal direction*	*Signal name*
1		x	Protective Ground
2	3	DTE-to-DCE	Transmitted Data (TxD)
3	2	DCE-to-DTE	Received Data (RxD)
4	7	DTE-to-DCE	Request To Send (RTS)
5	8	DCE-to-DTE	Clear To Send (CTS)
6	6	DCE-to-DTE	Data Set Ready (DSR)
7	5	x	Signal Ground (SG)
8	1	DCE-to-DTE	Received Line Signal Detector (Carrier Detect) (DCD)
20	4	DTE-to-DCE	Data Terminal Ready (DTR)
22	9	DCE-to-DTE	Ring Indicator (RI)

Let us see how two devices, one DTE and the other DCE, are connected through DB-25 connectors.

There are three classes of connections in a serial interface: a **signal common** line, two **data lines**, and five **handshake lines**.

Signal ground. (SG, Signal Ground). A signal ground (common) line must be connected between the DTE and DCE to provide a reference point for other signal voltages. DB-25 pin 7 of each device is designated as this signal common. Signal common should not be confused with protective ground, which is assigned to DB-25 pin 1. Protective ground is supposed to be "electrically bonded to the machine or equipment frame," but signal common "establishes the common reference potential for all interchange circuits" except protective ground.

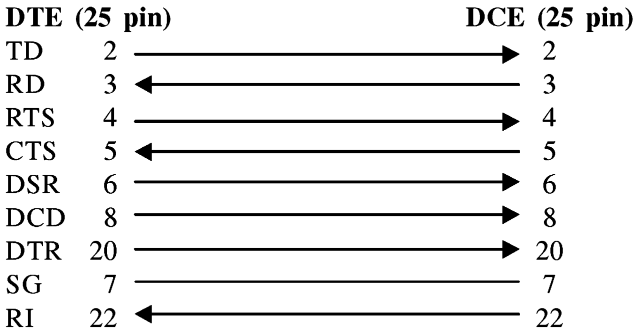
Data lines. (TxD, Transmit Data; RxD, Receive Data). A pair of these data lines are connected between the DTE and the DCE. According to RS 232, data on each of these lines flows in only *one* direction, so obviously a pair of lines is needed for two-way communication. By convention, the name of each data line is given with respect to the DTE. DB-25 pin 2 of the DTE is called TxD (Transmit Data), and DB-25 pin 3 of the DTE is called RxD (Receive Data). A DTE transmits on DB-25 pin 2 and receives on DB-25 pin 3. A DCE transmits on DB-25 pin 3 and receives on DB-25 pin 2. Further, the transmit pin of a serial device must have a negative voltage (MARKing) when no data is being transmitted. By probing DB-25 pins 2 and 3, we can identify the device whether it is DTE or DCE.

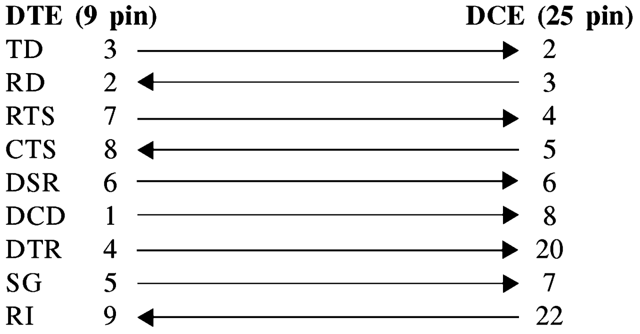
Handshake lines. (DTR, Data Terminal Ready; DSR, Data Set Ready; RTS, Request To Send; CTS, Clear To Send; and DCD, Data Carrier Detect). These lines provide information to each device about the status of the other device. The questions "Is the device turned on?" and "Is the device ready to receive data?" are answered by the state of the handshake lines. It is easy to see, from the following details, how these handshake functions can be used to regulate and control the flow of information between two serial devices.

DTR and DSR. Data Terminal Ready is a DTE output (DB-25 pin 20) that tells the connected DCE that the DTE is powered up and ready. Similarly, Data Set Ready is a DCE output (DB-25 pin 6) that tells the DTE that the DCE is powered up and ready. Ideally, both DTR and DSR have a positive voltage as long as each device has power. In practice, these connections are often omitted from a serial interface. The assumption is that both devices have been turned on before trying to send data between them. For example, a DTE may have an internal jumper that holds its DSR input high at all times, making it think that the connected DCE is on all the time.

RTS and CTS. Request To Send is a DTE output (DB-25 pin 4) that tells the connected DCE that the DTE is actually ready to receive data. Similarly, Clear To Send is a DCE output (DB-25 pin 5) that tells the connected DTE that the DCE is ready to receive data. Each line has positive voltage as long as the device is ready to receive. These two signals are, in practice, probably the most commonly used hardware handshake lines of the set. Connecting the RTS output of a DTE to the RTS input of a DCE will allow the DTE to control the flow of data from the DCE. If the DCE notices a negative voltage on the RTS line, then it will suspend data transmission to the DTE until it sees a positive voltage. The CTS line is used in a similar way to control the flow of data from the DTE to the DCE.

DCD. Data Carrier Detect is a DCE output (DB-25 pin 8) intended to inform the DTE that the DCE has a carrier. This line is reserved for use with modems, though we may sometimes find a non-modem device that uses this line to control data flow. The interconnection between DTE and DCE with DB-25 and DB-9 connectors is given as:





Maximum cable lengths

Cable length is one of the most discussed items in RS 232. As per the standard, the maximum cable length is 50 feet, or the cable length equal to a capacitance of 2500 pF. The latter rule is often forgotten. This means that using a cable with low capacitance allows us to span longer distances without going beyond the limitations of the standard. If, for example, Unshielded Twisted Pair (UTP) CAT-5 cable is used with a typical capacitance of 20 pF/m, the maximum allowed cable length is 125 metres.

The cable length mentioned in the standard allows maximum communication speed to occur. If speed is reduced by a factor 2 or 4, the maximum length increases dramatically. Table 14.2 gives the cable length at a particular baud rate.

Table 14.2 Cable length versus baud rate

Baud rate	*Maximum cable length* (ft)
19200	50
9600	500
4800	1000
2400	3000

Limitations of RS 232

RS 232 is simple, universal, well understood and supported everywhere. However, it has some serious shortcomings as an electrical interface. Firstly, the interface uses open signal lines with a common ground between the DTE and DCE. This is a reasonable assumption where a short cable connects a DTE and DCE in the same room, but with longer lines and connections between devices that may be on different electrical buses, this may not be true. Secondly, a signal on a single line is impossible to screen effectively for noise. By screening the entire cable, we can reduce the influence of outside noise, but internally-generated noise remains a problem. As the baud rate and line length increase, the effect of capacitance between the cables introduces serious crosstalk until a point is reached where the data itself is unreadable.

Crosstalk can be reduced by using low capacitance cable. Also, as it is the higher frequencies that are the problem, control of slew rate in the signal (i.e. making the signal more rounded, rather than square) decreases the crosstalk. The original specifications for RS 232 had no specification for maximum slew rate.

Wireless RS 232

Today there is a better way to create serial connections, and it does not require a connecting cable. An exciting new development, known as Bluetooth™ wireless technology, allows problematic RS 232 cables to be replaced with maintenance-free wireless communications. If we already use RS 232 cables then we can implement a wireless link without any modification to the equipment. The Wireless RS 232 Kit is available and can be used as a RS 232 cable replacement in all applications. Figure 14.12 shows the wireless interface concept.

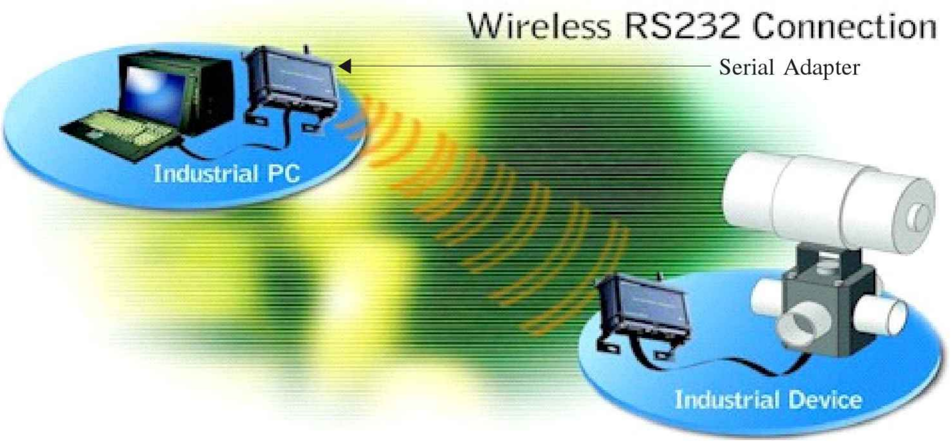


Figure 14.12 Wireless RS 232 connection concept.

Bluetooth™ wireless technology

Bluetooth wireless technology is a robust secure short-range wireless communication protocol developed as a cable replacement for computers, industrial devices, mobile telephones, hand-held devices, etc. Bluetooth radios operate in the free Industrial Scientific and Medical (ISM) Communications Band of 2.4 GHz to 2.5 GHz.

The Bluetooth concept was conceived in 1994 when Ericsson Mobile Communications began a study to examine alternatives for voice and data cabling. The Bluetooth Special Interest Group (SIG) was founded in 1996 with 5 members, Ericsson, Nokia, IBM, Intel and Toshiba. Microsoft, Lucent, 3COM and Motorola joined the original 5 core promoters in 1999. Others were invited to join as Bluetooth Adopters, the SIG now has over 2000 members from the fields of semiconductors, automotive and industrial, telecom, computing and peripherals, consumer, networking, and so forth.

Various feature of Bluetooth Technology are as follows:

- Operates in the 2.4 GHz ISM band, unlicensed (2.4000 –> 2.4835 GHz)
- Frequency hopping, 1600 hops/s
- Split into 79 channels (1 MHz/channel)
- GFSK modulated
- Data/voice
- Transmission rate up to 1 Mbit/s
- 10-100 metre range (Class dependant)
- Provides multiple levels of security

Wireless RS 232 units can be sealed making them ideal for use in industrial environments. Communication is possible over distances up to 30 metres.

14.3.4 RS 422

When communicating at high data rates, or over long distances in real world environments, single-ended methods, like RS 232, are often inadequate. Differential data transmission (balanced differential signal) offers superior performance in most applications. Differential signals can help nullify the effects of ground shifts and induced noise signals that can appear as common mode voltages on a network.

By title, the RS 422 standard is TIA/EIA 422 B, “Electrical Characteristics of Balanced Voltage Digital Interface Circuits” by the Telecommunications Industry Association (in association with the EIA). RS 422 (differential) was designed for greater distances and higher baud rates than RS 232. It is almost similar to RS 485, the main differences being the rise times and voltage characteristics of the waveform. RS 422 generally allows cable lengths up to 1.2 kilometres at up to 100 Kbps. At 10 Mbps, cable lengths are limited to around 10 metres (Figure 14.13). In the presence of cable imbalance or high common mode noise levels, cable lengths may be further reduced in order to maintain a desired signaling rate.

RS 422 is full-duplex, i.e. data can flow in both directions simultaneously and therefore uses two separate twisted pairs. RS 422 is often used simply for extending RS 232 cables.

RS 422 is also specified for multi-drop (party-line) applications where only one driver is connected to, and transmits on, a bus of up to 10 receivers. While a multi-drop type application has many desirable advantages, RS 422 devices cannot be used to construct a truly multi-point network. A true multi-point network consists of multiple drivers and receivers connected on a single bus, where any node can transmit or receive data.

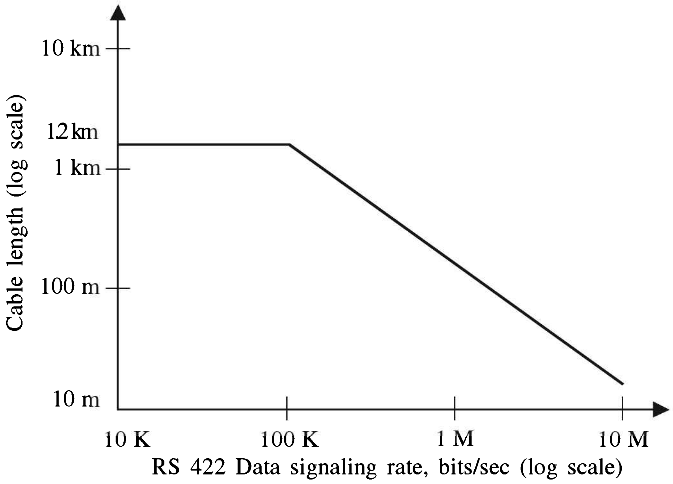


Figure 14.13 Cable length versus RS 422 data rate.

"Quasi" multi-drop networks (4-wire) are often constructed using RS 422 devices. These networks are often used in a half-duplex mode, where a single master in a system sends a command to one of several *slave* devices on a network. Typically one device (node) is addressed by the host computer and a response is received from that device. Systems of this type (4-wire, half-duplex) are often constructed to avoid *data collision* (bus contention) problems on a multi-drop network (more about solving this problem on a two-wire network in a moment).

Balanced line interface

In RS 422, each signal is sent differentially over two separate wires, denoted as A and B. This is called *balanced line interface*. Figure 14.14 shows how the transmission is carried from the driver (transmitter) to the receiver. Differential signals are produced by the drivers and translated back to single ended-signals by the receivers.

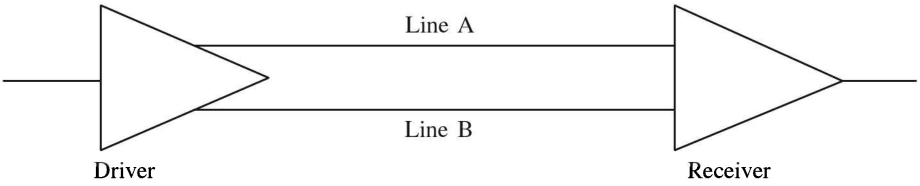


Figure 14.14 RS 422 transmission in balanced mode.

The higher transmission rate is possible because differential lines are terminated by resistors so they act as transmission lines instead of simply open wires. Further, a common-mode electrical noise induced in the two lines is rejected by the differential line receiver.

A logic 0 or mark is indicated by B signal line being more positive than the A signal line. A logic 1 or space is indicated by A signal line being more positive than the B signal line. The voltage difference between the two lines must be greater than 0.4 V, but less than 12 V. So,

$$V_A - V_B < -0.2\text{ V} = \text{logic } 0$$

$$V_A - V_B > +0.2\text{ V} = \text{logic } 1$$

The common mode voltage on the signal lines must be in the range of –7 V to +7 V.

If lines A and B are close together, they will be affected almost identically by external electromagnetic noise. If the lines are also twisted together, then neither line is permanently closer to a noise source than the other. Hence the well known *twisted pair* is extremely effective in eliminating noise from the signal. To identify lines A and B, we can measure the voltages when no communication is taking place. B should be more positive than A.

Converters are available to interface RS 232 with RS 422 and vice versa. RS 422 does not specify connector pin numbers or handshake signals like RS 232.

14.3.5 RS 423

RS 423 is another single-ended specification with enhanced operation over RS 232. This standard specifies a low impedance single-ended signal instead of differential signal of RS 422. The low impedance signal can be sent over 50 Ω coaxial cable and terminated at the receiving end to prevent reflections.

We can assume the same arrangement as shown in the Figure 14.14, where line B is at ground potential. A logic 0 or mark is indicated by B signal line, being 4–6 V negative than the A signal line. A logic 1 or space is indicated by A signal line, being 4–6 V positive than the B signal line.

RS 423 does not specify connector pin numbers or handshake signals like RS 232.

14.3.6 RS 485

RS 485 is another serial data transmission method. Officially, it is EIA 485, or “Standard for Electrical Characteristics of Generators and Receivers for Use in Balanced Digital Multipoint Systems” by the Electronics Industry Association (EIA). This is a specialized interface that is very common in the data acquisition applications.

RS 485 is half-duplex. It exists in two varieties—2-wire (which uses a single twisted pair) and 4-wire (which uses two twisted pairs like RS 422). RS 485 systems are usually “Master/Slave”; each Slave device has a unique address and it responds only to a correctly addressed message (a *poll*) from the Master. A Slave never initiates a dialogue. In a 2-wire system, all devices (including the Master) must have tri-state capability. In a 4-wire system, the Slaves must have tri-state capability but the Master does not need it (i.e. it can be an RS 422 device) because it drives a bus on which all other devices merely listen.

RS 485 allows multiple drivers and receivers (32 drivers and 32 receivers). It specifies cable characteristics in terms of signaling speeds and lengths or total capacitance. A typical cable is a shielded twisted copper pair, which is adequate for the typical signaling rate of 10 Mbps. Maximum cable length can be as much as 4000 feet because of the differential voltage transmission system used. With very long cables, cable capacitance becomes the dominant factor in the power consumption, much more significant than transmission line effects. Table 14.3 lists a very conservative value of the capacitance at different baud rates. Cable capacitance is typically 100 pF/m.

Table 14.3 Baud rate versus cable capacitance

Baud rate	*Max total cable capacitance* (pF)	*Baud rate*	*Max total cable capacitance* (pF)
1200	250,000	19,200	15,000
2400	120,000	38,400	7500
4800	60,000	57,600	5000
9600	30,000	115,200	2500

One example of a simple RS 485 serial network is a series of recorders connected over an RS 485 link to a PC that receives data acquired by each recorder. A plug-in card is installed in each recorder. Each network card is daisy-chained to the others over a series of shielded twisted pair cables that ultimately terminate on a network interface card in the PC. There is no real need to know and understand the network layers in this arrangement, except to understand the limitations on RS 485 (distance, shielding, data rate, etc.).

We note that both RS 422 and RS 485 use a twisted-pair wire for each signal and the same differential interface with identical voltages. The main difference between them is as follows: RS 422 has no tri-state capability (its driver is always enabled) and it is therefore usable only in point-to-point communications (although an RS 422 device can act as a Master on a 4-wire RS 485 system). RS 485 has tri-state capability and can therefore be used in multidrop systems.

Table 14.4 compares the specifications of RS 232, RS 422, RS 423 and RS 485.

Table 14.4 Specifications of RS 232, RS 422, RS 423 and RS 485

Specifications		*RS 232*	*RS 423*	*RS 422*	*RS 485*
Mode of operation		Single ended	Single ended	Differential	Differential
Total number of drivers and receivers on one line		1 Driver 1 Receiver	1 Driver 10 Receiver	1 Driver 10 Receiver	1 Driver 32 Receiver
Maximum cable length		50 ft	4000 ft	4000 ft	4000 ft
Maximum data rate		20 Kbps	100 Kbps	10 Mbps	10 Mbps
Maximum driver output voltage		±25 V	±6 V	–0.25 V to +6 V	–7 V to 12 V
Driver output signal level (loaded min.)	Loaded	±5 V to ±15 V	±3.6 V	±2.0 V	±1.5 V
Driver output signal level (unloaded max)	Unloaded	±25 V	± 6V	±6 V	±6 V
Driver load impedance (ohms)		3 kΩ to 7 kΩ	≥ 450 Ω	100 Ω	54 Ω
Max. driver current in high *Z* state	Power on	N/A	N/A	N/A	±100 μA
Max. driver current in high *Z* state	Power off	±6 mA @ ±2 V	±100 μA	±100 μA	±100 μA
Slew rate (max.)		30 V/μs	Adjustable	N/A	N/A
Receiver input voltage range		±15 V	±12 V	–10 V to +10 V	–7 V to +12 V
Receiver input sensitivity		±3 V	±200 mV	±200 mV	±200 mV
Receiver input resistance (ohms)		3 kΩ to 7 kΩ	4 kΩ min.	4 kΩ min.	≥ 12 kΩ

14.3.7 IEEE 488 (GPIB)

All the standards described above are used in serial communication. There is one parallel interface, known as the General Purpose Interface Bus (GPIB), or IEEE 488. It was designed by Hewlett-Packard in 1965 as the Hewlett-Packard Interface Bus (HP-IB) to connect programmable instruments to the computers. Because of its high transfer rates (nominally 1 Mbytes/s), this interface bus quickly gained popularity. It was later accepted as IEEE Standard 488 (1975), and has evolved to ANSI/IEEE Standard 488.1 (1987). Today, the name General Purpose Interface Bus (GPIB) is more widely used than HP-IB. ANSI/IEEE 488.2 (1987) strengthened the original standard by defining precisely how controllers and instruments communicate. Standard Commands for Programmable Instruments (SCPI) took the command structures defined in IEEE 488.2 and created a single, comprehensive programming command set that is used with any SCPI instrument. Almost any instrument can be used with the GPIB specification by adding the interface to the instrument. The standard defines the signals passing into the interface from the GPIB bus and from the instrument to the GPIB bus. The instrument does not have complete control over the interface. But an Active Controller performs the bus control functions for all the bus instruments.

GPIB devices communicate with the other GPIB devices by sending device-dependent messages and interface messages through the interface system.

- Device-dependent messages, often called data or data messages, contain device-specific information, such as programming instructions, measurement results, machine status, and data files.
- Interface messages manage the bus. Usually called commands or command messages, interface messages perform such functions as initializing the bus, addressing and unaddressing devices, and setting device modes for remote or local programming.

Talkers, listeners, and controllers

GPIB devices can be Talkers, Listeners, and/or Controllers (see Figure 14.15). A talker sends data messages to one or more listeners, which receive the data. The Controller manages the

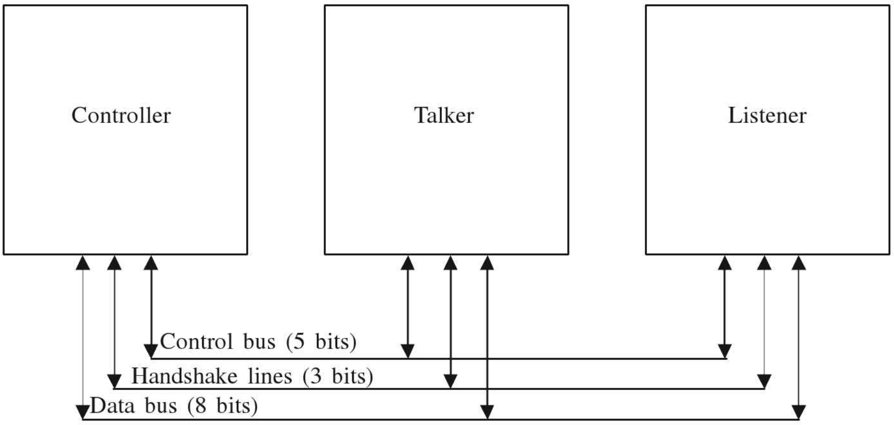


Figure 14.15 Devices connections.

flow of information on the GPIB by sending commands to all devices. A digital voltmeter, for example, is a talker and is also a listener.

The GPIB is like an ordinary computer bus, except that a computer has its circuit cards interconnected via a backplane, while the GPIB has stand-alone devices interconnected by standard cables.

The role of the GPIB Controller is like the controller in the switching centre of a telephone exchange. The switching centre (Controller) monitors the communications network (GPIB). When the centre (Controller) notices that a party (device) wants to make a call (send a data message), it connects the caller (talker) to the receiver (listener).

The Controller usually addresses (or enables) a talker and a listener before the talker can send its message to the listener. After the message is transmitted, the controller may address other talkers and listeners. Some GPIB configurations do not require a controller. For example, a device that is always a talker called a talk-only device, is connected to one or more listen-only devices.

A controller is necessary when the active or addressed talker or listener must be changed. The function of the controller is usually handled by a computer.

The standard allows a maximum of 15 devices to be connected on the same bus. A minimum system consists of one controller and one talker or listener. It is possible to have several controllers on the bus but only one may be active at any given time. The Active Controller (controller-in-charge, CIC) may pass control to another controller which in turn can pass it back or on to another controller. The controller can set up a talker and a group of listeners so that it is possible to send data between groups of devices as well.

Connector and signal lines

The IEEE 488 bus is a multi-drop, parallel interface with 24 lines accessed by all devices. The lines are grouped into data lines, handshake lines, bus management lines, and ground lines. Communication is digital, and messages are sent one byte at a time. The connector is a 24-pin connector; devices on the bus use female receptacles while interconnecting cables have mating male plugs. A typical cable will have male and female connectors to allow daisy-chaining between devices. The cable is 24 conductor shielded cable. It provides a high data signaling rate, up to 1 Mbps, but cable is limited in length. The total bus length permitted is 20 metres, with no more than 4 metres between devices. Figure 14.16 shows the connector and pin configuration. The function of all the signal lines is explained in the following subsections.

Data bus

The eight data lines DIO1 through DIO8 are used to transfer addresses, and control information and data. The state of the Attention (ATN) line determines whether the information is data or commands. The formats for addresses and control bytes are defined by the standard. Data formats are undefined and may be 7-bit ASCII or ISO code set, in which case the eighth bit, DIO8, is either unused or used for parity. DIO1 is the least significant bit (note that this will correspond to bit 0 on most computers).

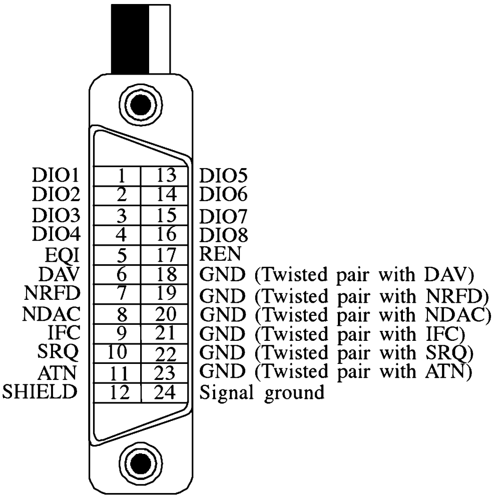

DIO1
DIO2
DIO3
DIO4
EQI
DAV
NRFD
NDAC
IFC
SRQ
ATN
SHIELD
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
DIO5
DIO6
DIO7
DIO8
REN
GND (Twisted pair with DAV)
GND (Twisted pair with NRFD)
GND (Twisted pair with NDAC)
GND (Twisted pair with IFC)
GND (Twisted pair with SRQ)
GND (Twisted pair with ATN)
Signal ground

Figure 14.16 IEEE 488 connector and pin configuration.

Handshake lines

The three handshake lines (NRFD, NDAC, DAV) asynchronously control the transfer of message bytes among the devices and form the method for acknowledging the transfer of data. This handshaking process guarantees that the bytes on the data lines are sent and received without any transmission errors and is one of the unique features of the GPIB bus. The process is called a 3-wire interlocked handshake.

The NRFD (Not Ready for Data) handshake line is asserted by a listener to indicate it is not yet ready for the next data or control byte. Note that the controller will not see NRFD released (i.e. ready for data) until all devices have released it.

The NDAC (Not Data Accepted) handshake line is asserted by a listener to indicate it has not yet accepted the data or control byte on the data lines. Note that the controller will not see NDAC released (i.e. data accepted) until all devices have released it.

The DAV (Data Valid) handshake line is asserted by the talker to indicate that a data or control byte has been placed on the data lines and has had the minimum specified stabilizing time. The byte can now be safely accepted by the devices.

Handshaking

In the handshaking process, when the controller or a talker wishes to transmit data on the bus, it sets the DAV line HIGH (data not valid), and checks to see that the NRFD and NDAC lines are both LOW, and then it puts the data on the data lines.

When all the devices that can receive the data are ready, each releases its NRFD (not ready for data) line. When the last receiver releases NRFD, and it goes HIGH, the controller or talker takes DAV LOW indicating that valid data is now on the bus.

In response each receiver takes NRFD LOW, again to indicate that it is busy and releases NDAC (not data accepted) when it has received the data. When the last receiver has accepted the data, NDAC will go HIGH and the controller or talker can set DAV HIGH again to transmit the next byte of data.

Note that if after setting the DAV line HIGH, the controller or Talker senses that both NRFD and NDAC are HIGH, an error will occur. Also if any device fails to perform its part of the handshake and releases either NDAC or NRFD, data cannot be transmitted over the bus. Eventually a time-out error will be generated.

The speed of the data transfer is controlled by the response of the slowest device on the bus, for this reason it is difficult to estimate data transfer rates on the GPIB bus as they are always device dependent.

Control bus

The five control or interface management lines (ATN, EOI, IFC, REN, SRQ) manage the flow of control and data bytes across the interface.

The ATN (Attention) signal is asserted by the controller to indicate that it is placing an address or control byte on the data bus. ATN is released to allow the assigned talker to place status or data on the data bus. The controller regains control by reasserting ATN; this is normally done synchronously with the handshake to avoid confusion between control and data bytes.

The EOI (End or Identify) signal has two uses. A talker may assert EOI simultaneously with the last byte of data to indicate end-of-data. The controller may assert EOI along with ATN to initiate a parallel poll. Although many devices do not use parallel poll, all devices should use EOI to end transfers (many currently available ones do not).

The IFC (Interface Clear) signal is asserted only by the System Controller in order to initialize all device interfaces to a known state. After releasing IFC, the System Controller is the Active Controller.

The REN (Remote Enable) signal is asserted only by the System Controller. Its assertion does not place devices into remote control mode; REN only enables a device to go into remote mode when addressed to listen. When in remote mode, a device should ignore its local front panel controls.

The SRQ (Service Request) line is like an interrupt: it may be asserted by any device to request the controller to take some action. The controller must determine which device is asserting SRQ by conducting a serial poll. The requesting device releases SRQ when it is polled.

All lines in the GPIB are tri-state except for SQR, NRFD, and NDAC which are open-collector. The standard bus termination is a 3 kΩ resistor connected to 5 V in series with a 6.2 kΩ resistor to ground—all values having a 5% tolerance.

Data bus transfer timing

Figure 14.17 shows the timing diagram illustrating how data is read.

Device Addresses

The IEEE 488 standard allows up to 15 devices to be interconnected on one bus. Each device

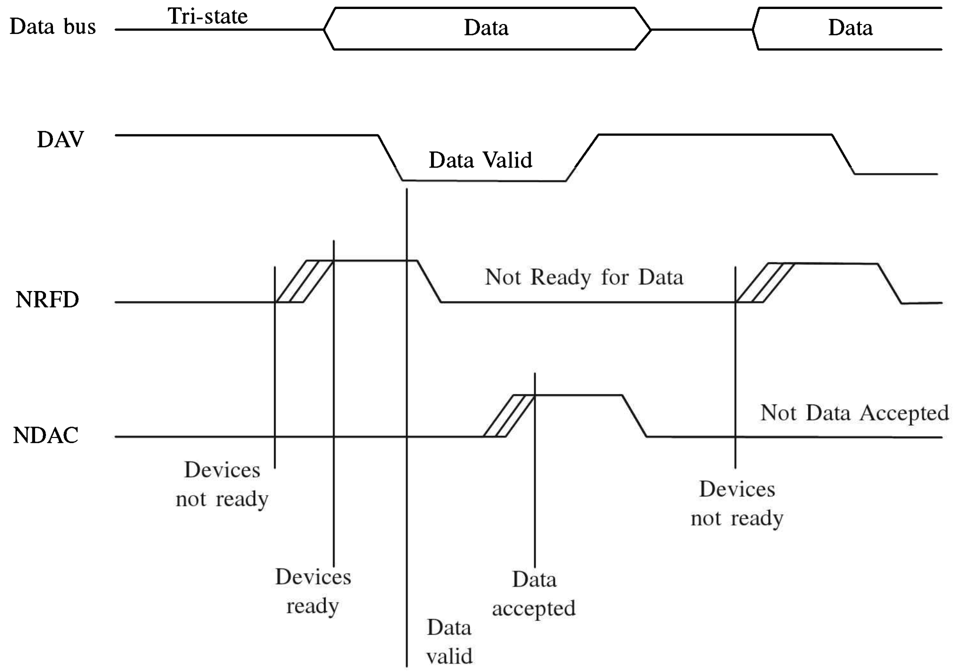


Figure 14.17 Timing diagram.

is assigned a unique primary address, ranging from 0-30, by setting the address switches on the device. A secondary address may also be specified, ranging from 0-30.

The standard also allows for identification of the devices on the bus. Each device should have a string of 1 or 2 letters placed somewhere on the body of the device (near or on the GPIB connector). These letters signify the capabilities of the device on the GPIB bus.

C	Controller
T	Talker
L	Listener
AH	Acceptor Handshake
SH	Source Handshake
DC	Device Clear
DT	Device Trigger
RL	Remote Local
PP	Parallel Poll
TE	Talker Extended
LE	Listener Extended

An example of an IEEE 488 implementation is a measurement system designed to assess the performance of a chemistry sample sink. The sink performs sample conditioning (pressure, flow, and temperature control) and chemical analysis (pH, dissolved oxygen, and conductivity) on water samples. The sink is instrumented with pressure sensors, resistance temperature detectors (RTDs), thermocouples, and reference junctions. A 30-point scanner is used to multiplex data from all of the sensors. The scanner is connected to a desktop or laptop PC using the GPIB interface. Data is acquired, stored, displayed, and reduced using application programs on the PC, efficiently and reliably under IEEE 488.

Physical connection

Devices can be connected in either a linear (Figure 14.18), star (Figure 14.19) or combination configuration.

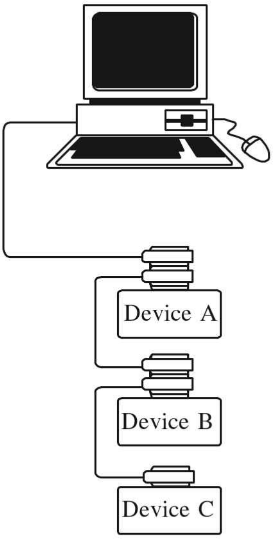


Figure 14.18 Daisy chaining or linear configuration.

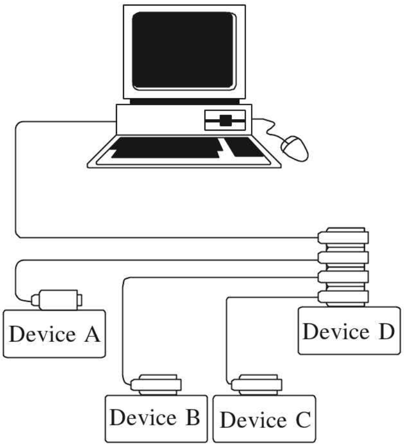


Figure 14.19 Star configuration.

The media used to connect the devices is usually a set of copper wires. Unshielded twisted pair (UTP) cable is the most affordable. It is lightweight, easy to pull, easy to terminate, and uses less cable tray space than the shielded twisted pair (STP). However, it is more susceptible to electromagnetic interference (EMI).

STP is heavier and more difficult to manufacture, but it can greatly improve the signaling rate in a given transmission scheme. Twisting provides cancellation of magnetically induced fields and currents on a pair of conductors. Magnetic fields arise around other heavy current-carrying conductors and around large electric motors. Various grades of copper cables are available, with Grade 5 being the best and most expensive. Grade 5 copper, appropriate for use in 100-Mbps applications, has more twists per inch than lower grades. More twists per inch means more linear feet of copper wire used to make up a cable run, and more copper means more cost.

Shielding provides a means to reflect or absorb electric fields that are present around cables. Shielding comes in a variety of forms from copper braiding or copper meshes to aluminized mylar tape wrapped around each conductor and again around the twisted pair.

Fiber optics is being used more often as user applications demand higher and higher bandwidths. There are three basic fiber optic cables available: multimode step index, multimode graded index, and single mode. Multimode fibers usually are driven by LEDs at each end of the cable, while single mode fibers usually are driven by lasers. Single-mode fibers can achieve much higher bandwidths than multimode fibers, but are thinner (10 microns) and physically weaker than multimode. Equipment costs for transmitting and receiving single-mode fiber signals are much higher (at least four times) than those for multimode signals.

One distinct advantage of fiber optic cables is noise immunity. Fiber optic cables can be routed indiscriminately through high noise areas with impunity, although fire ratings should be observed. Cables that pass through multiple spaces in a plant should be rated for heating/ventilation/air conditioning (HVAC) plenums where they can withstand fires.

Electrical characteristics

The bus uses standard TTL level negative logic. When NRFD is true, for example, it is a TTL low level, and when NRFD is false, it is a TTL high level. The I/O characteristics are given in Table 14.5.

Table 14.5 I/O characteristics of GPIB bus

Signal type	*Value*
Input voltage	V_{IH} = 3.4 volts typical, 2.4 volts minimum
	V_{IL} = 0.22 volts typical, 0.4 volts maximum
Input current	I_{IH} = 2.5 mA, maximum
	I_{IL} = –3.2 mA, maximum
Output voltage	V_{OH} = 3.4 volts typical, 2.5 volts minimum
	V_{L} = 0.22 volts typical, 0.5 volts maximum
Output current	I_{OH} = –5.2 mA, maximum
	I_{OL} = 48 mA, maximum

IEEE 488.2

The IEEE 488.1 standard, as discussed above, greatly simplifies the interconnection of programmable instruments by clearly defining mechanical, hardware, and electrical protocol specifications. This standard does not address data formats, status reporting, message exchange protocol, common configuration commands, or device specific commands.

The IEEE 488.2 standard enhances and strengthens the IEEE 488.1 standard by specifying data formats, status reporting, error handling, controller functionality, and common instruments commands. It focuses mainly on the software protocol issues and thus maintains compatibility with the hardware-oriented IEEE 488.1 standard. IEEE 488.2 systems tend to be more compatible and reliable. The IEEE 488.2 standard focuses mainly on the software protocol issues and thus maintains compatibility with the hardware-oriented IEEE 488.1 standard.

SCPI (Standard Commands for Programmable Instruments), built on the IEEE 488.2 standard, simplifies the programming task by defining a single comprehensive command set for programmable instrumentation, regardless of the type or manufacturer. SCPI systems are much easier to program and maintain. In many cases, we can interchange or upgrade instruments without having to change the test program. The combination of SCPI and IEEE

488.2 offers significant productivity gains, and finally, delivers as sound a software standard as IEEE 488.1 did a hardware standard. The scope of each of the IEEE 488, IEEE 488.2, and SCPI standards is shown in Figure 14.20.

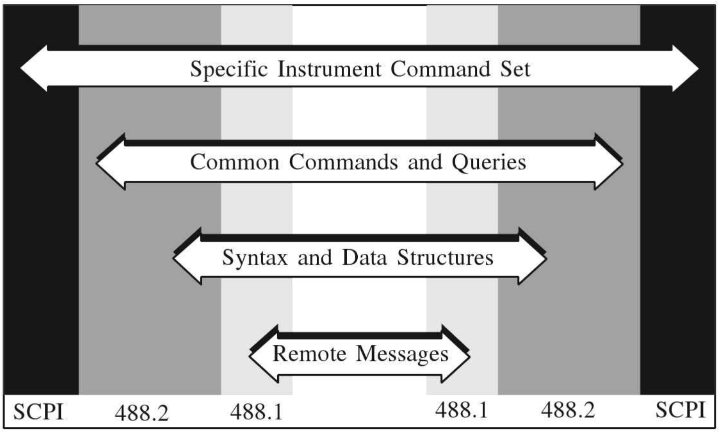

Specific Instrument Command Set
Common Commands and Queries
Syntax and Data Structures
Remote Messages
SCPI
488.2
488.1
488.1
488.2
SCPI

Figure 14.20 Evolution of GPIB instrumentation standards.

IEEE 488.2 (1987) encourages a new level of growth and acceptance of the IEEE 488 bus or GPIB by addressing problems that had arisen from the original IEEE 488 standard. IEEE 488.2 stays compatible with the existing IEEE 488.1 standard. The overriding concept used in the IEEE 488.2 specification for the communication between the controllers and instruments is that of *precise talking* and *forgiving listening*. In other words, IEEE 488.2 exactly defines how both IEEE 488.2 controllers and IEEE 488.2 instruments talk so that a completely IEEE 488.2 compatible system can be highly reliable and efficient. The standard also requires that IEEE 488.2 devices be able to work with the existing IEEE 488.1 devices by accepting a wide range of commands and data formats as a listener. We obtain the true benefits of IEEE 488.2 when we have a completely IEEE 488.2 compatible system.

IEEE 488.2 control sequences and protocols

The IEEE 488.2 standard defines the control sequences that specify the exact IEEE 488.1 messages that are sent from the controller as well as it defines the ordering of multiple messages. IEEE 488.2 has defined 15 required control sequences and four optional control sequences, as shown in Table 14.6.

Table 14.6 Control sequences of IEEE 488.2

Description	*Control sequence*	*Compliance*
Send ATN-true commands	SEND COMMAND	Mandatory
Send address to send data	SEND SETUP	Mandatory
Send ATN-false data	SEND DATA BYTES	Mandatory
Send a program message	SEND	Mandatory
Send address to receive data	RECEIVE SETUP	Mandatory
Receive ATN-false data	RECEIVE/RESPONSE MESSAGE	Mandatory
Receive a response message	RECEIVE	Mandatory
Pulse IFC line	SEND IFC	Mandatory
Place device in DCAS	DEVICE CLEAR	Mandatory
Place devices in local state	ENABLE LOCAL CONTROLS	Mandatory
Place devices in remote state	ENABLE REMOTE	Mandatory
Place devices in remote with local lockout state	SET RWLS	Mandatory
Place devices in local lockout state	SEND LLO	Mandatory
Read IEEE 488.1 status byte	READ STATUS BYTE	Mandatory
Send group execution trigger (GET) message	TRIGGER	Mandatory
Give control to another device	PASS CONTROL	Optional
Conduct a parallel poll	PERFORM PARALLEL POLL	Optional
Configure devices' parallel poll responses	PARALLEL POLL CONFIGURE	Optional
Disable devices' parallel poll capability	PARALLEL POLL UNCONFIGURE	Optional

Protocols are high-level routines that combine a number of control sequences to perform common test system operations. IEEE 488.2 defines two required protocols and six optional protocols, as shown in Table 14.7.

Table 14.7 IEEE 488.2 protocols

Keyword	*Name*	*Compliance*
RESET	Reset System	Mandatory
FINDRQS	Find Device Requesting Service	Optional
ALLSPOLL	Serial Poll All Devices	Mandatory
PASSCTL	Pass Control	Optional
REQUESTCTL	Request Control	Optional
FINDLSTN	Find Listeners	Optional
SETADD	Set Address	Optional, but requires FINDLSTN
TESTSYS	Self-Test System	Optional

HS 488

Innovations in technology continue to improve the computers and the instrumentation that we use to implement the GPIB-based test and measurement applications. Faster microprocessors and memory, PCI bus DMA, and advancements in digitizer technology contribute to reducing the time to capture and analyze large volumes of test data. As technology continues to evolve, the speed at which data can be transferred from the instrumentation to the host computer becomes important. Anticipating the need for a higher-speed data transfer mechanism for GPIB, National Instruments has created HS 488, which delivers a standard method for increasing the maximum GPIB data transfer rate up to 8 Mbytes/s while maintaining compatibility with existing GPIB applications and instrumentation.

HS 488, a higher-speed GPIB transfer protocol, scales the maximum data transfer rate of ANSI/IEEE Standard 488.1 (1987) up to 8 Mbytes/s by removing delays in the 3-wire IEEE 488.1 handshake. HS 488 is implemented in hardware as a feature of the TNT 4882C GPIB controller ASIC, and thus imposes no additional software overhead during HS 488 transfers. Because HS 488 is implemented at the hardware level, the application code requires no modification to take advantage of high-speed GPIB. Using the HS 488 protocol, the GPIB controller hardware can automatically detect compatible devices capable of using the HS 488 handshake to transfer data. If the controller does not detect an HS 488 capable device, it automatically defaults to the standard IEEE 488.1 3-wire handshake to complete the data transfer. The HS 488 protocol employs the same proven, high-speed, data streaming techniques used in VME, PCI, and Fast SCSI.

To ensure compatibility with the existing GPIB devices, the HS 488 protocol has been developed as a superset of the IEEE 488 standard; thus, we can mix IEEE 488.1, IEEE 488.2, and HS 488 devices in the same system and continue using existing cables and cabling configurations. HS 488 enhances the IEEE 488 standard, a foundation technology for test and measurement applications, by delivering additional performance and headroom to developers who want faster data transfers across GPIB.

Handshake

The standard IEEE 488.1 3-wire handshake (shown in Figure 14.21) requires the listener to unassert Not Ready for Data (NRFD), the talker to assert the Data Valid (DAV) signal to indicate to the listener that a data byte is available, and for the listener to unassert the Not Data Accepted (NDAC) signal when it has accepted that byte. A byte cannot transfer in less than the time it takes for the following events to occur: NRFD to propagate to the talker, DAV signal to propagate to all listeners, the listeners to accept the byte and assert NDAC, the NDAC signal to propagate back to the talker, and the talker to allow a settling time (T1) before asserting DAV again.

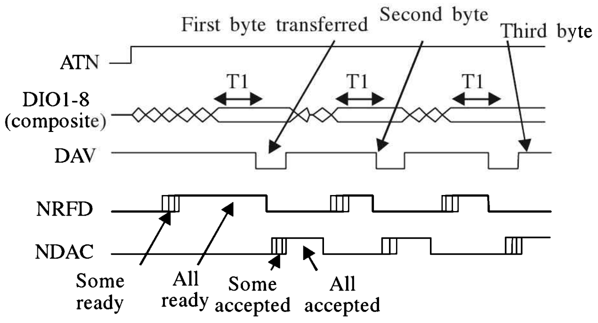


Figure 14.21 IEEE 488.1 handshake.

HS 488 increases the system throughput by removing propagation delays associated with the 3-wire handshake. To enable the HS 488 handshake, the talker pulses the NRFD signal line after the controller addresses all listeners. If the listener is HS 488 capable, then the transfer occurs using the HS 488 handshake (shown in Figure 14.22). Once HS 488 is enabled, the talker places a byte on the GPIB DIO lines, waits for a preprogrammed settling time, asserts DAV, waits for a preprogrammed hold time, unasserts DAV, and drives the next data byte on the DIO lines. The listener keeps NDAC unasserted and must accept the byte within the specified hold time. A byte must transfer in the time set by the settling time and hold time, without waiting for any signals to propagate along the GPIB cable.

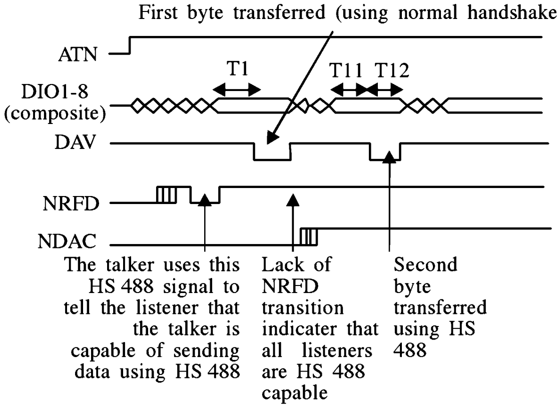

First byte transferred (using normal handshake)
ATN
T1
T11 T12
DIO1-8
(composite)
DAV
NRFD
NDAC
The talker uses this HS 488 signal to tell the listener that the talker is capable of sending data using HS 488
Lack of NRFD transition indicater that all listeners are HS 488 capable
Second byte transferred using HS 488

Figure 14.22 HS 488 handshake.

HS 488 data transfer flow control

The listener may assert NDAC to temporarily prevent more bytes from being transmitted, or assert NRFD to force the talker to use the 3-wire handshake. Through these methods, the listener can limit the average transfer rate. However, the listener must have an input buffer that can accept short bursts of data at the maximum rate, because by the time NDAC or NRFD propagates back to the talker, the talker may have already sent another byte.

The required settling and hold times are user configurable, depending on the total length of cable and number of devices in the system. Between two devices and 2 m of cable, HS 488 can transfer data up to 8 Mbytes/s. For a fully loaded system with 15 devices and 15 m of cable, HS 488 transfer rates can reach 1.5 Mbytes/s.

HS 488 controllers always use the standard IEEE 488.1 3-wire handshake to transfer GPIB commands (bytes with Attention (ATN) asserted).

System configuration effects on HS 488

The HS 488 protocol increases maximum transfer rates for GPIB reads and writes up to 8 Mbytes/s, depending on the system configuration. Maximum data transfer rates can be limited by the host computer and GPIB system setup. For PC-compatible computers with PCI bus, the maximum transfer rate obtainable is 8 Mbytes/s. For PC-compatible computers with an

ISA bus it is 2 Mbytes/s. The same IEEE 488 cabling constraints for a 350 ns T1 delay also apply to HS 488. As we increase the length of cable in the GPIB configuration, the maximum data transfer rate using HS 488 decreases. For example, two HS 488 devices connected by 2 m of cable can transfer data faster than four HS 488 devices connected by 4 m of cable. Figure 14.23 illustrates that with HS 488, we can always realize significant performance improvement over the 1 Mbyte/s, IEEE 488.1 3-wire handshake regardless of the cable length.

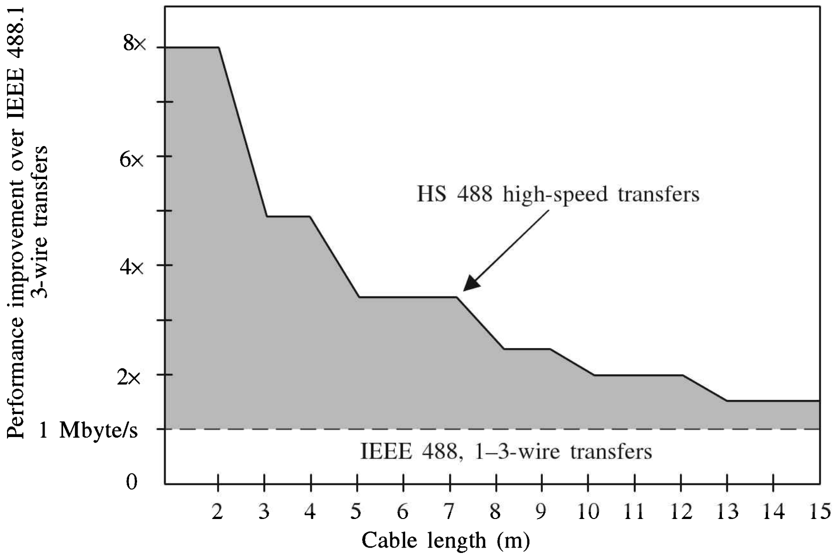


Figure 14.23 HS488 performance improvement over IEEE 488.1 3-wire transfers.

If we specify an HS 488 cable length that is much smaller than what we actually use, the transferred data could become corrupted. If we specify a cable length longer than we actually use, the data is transferred successfully, but more slowly than if we specified the correct cable length.

HS 488 configuration messages

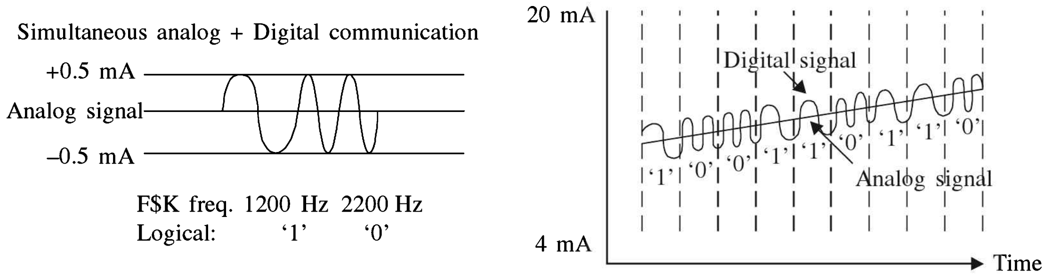
The controller-in-charge sends out the HS 488 command bytes (interface messages) to enable HS 488 on all HS 488 capable devices. These messages are sent only once to configure the system for HS 488. The HS 488 configuration message is made up of two GPIB command bytes. The first byte, the Configure Enable (CFE) message (hex 1F), places all HS 488 devices into their configuration mode. Non-HS 488 devices ignore this message. The second byte is a GPIB secondary command that indicates the number of metres of cable in the system. It is called the Configure (CFGn) message. Because HS 488 can operate only with cable lengths of 1 to 15 m, only CFGn values of 1 through 15 (hex 61 through 6F) are valid. If the cable length is configured properly in the GPIB Configuration Utility, we can determine how many metres of cable are in our system by calling ibask (option IbaHSCableLength) in our application program.

14.3.8 HART

For many years, the field communication standard for process automation equipment has been a milliamp (mA) analog current signal. The milliamp current signal varies within a range of 4–20 mA in proportion to the process variable being represented. In typical applications a signal of 4 mA will correspond to the lower limit (0%) of the calibrated range and 20 mA will correspond to the upper limit (100%) of the calibrated range. If the system is calibrated for 0–100 PSI, then an analog current signal of 12 mA (50% of range) will correspond to a pressure of 50 psi. Virtually, all installed systems use this international standard for communicating process variable information between process automation equipment.

HART Field Communications Protocol extends this 4–20 mA standard to enhance communication with smart field instruments. The HART protocol was designed specifically for use with intelligent measurement and control instruments which traditionally communicate using 4–20 mA analog signals. HART preserves the 4–20 mA signal and enables two-way digital communications to occur without disturbing the integrity of the 4–20 mA signal. Unlike other digital communication technologies, the HART protocol maintains compatibility with the existing 4–20 mA systems, and in doing so, provides users with a uniquely backward compatible solution. HART Communication Protocol is well-established as the *de facto* industry standard for digitally enhanced 4–20 mA field communication.

HART is an acronym for Highway Addressable Remote Transducer. The HART protocol makes use of the Bell 202 Frequency Shift Keying (FSK) standard to superimpose digital communication signals at a low level on top of the 4–20 mA as shown in Figure 14.24(a) and (b). This enables two-way field communication to take place and makes it possible for additional information beyond just the normal process variable to be communicated to/from a smart field instrument. The HART protocol communicates at 1200 bps without interrupting the 4–20 mA signal and allows a host application (master) to get two or more digital updates per second from a field device. As the digital FSK signal is phase continuous, there is no interference with the 4–20 mA signal.



(a) (b)

Figure 14.24 HART communication technique.

HART is a master/slave protocol which means that a field (slave) device only speaks when spoken to by a master. The HART protocol can be used in various modes for communicating information to/from smart field instruments and central control or monitoring systems. HART provides for up to two masters (primary and secondary) as shown in Figure 14.25. This

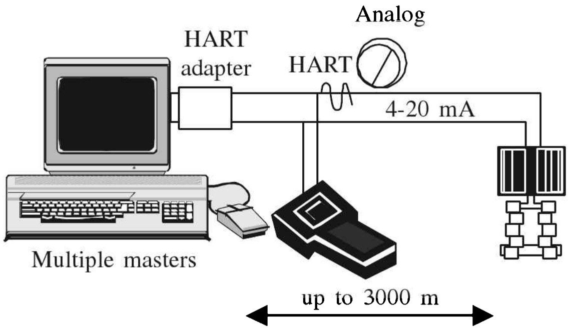
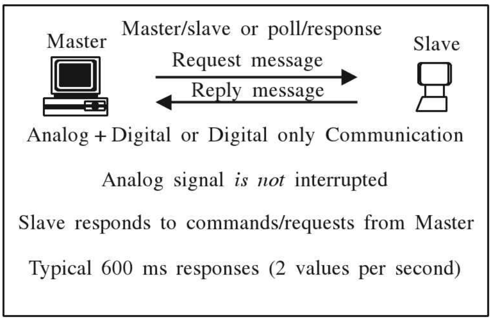


Figure 14.25 Two masters communicating with the slave.

allows secondary masters such as handheld communicators to be used without interfering with communications to/from the primary master, i.e. control/monitoring system. The most commonly employed HART communication mode is master/slave communication of digital information simultaneous with transmission of the 4–20 mA signal as shown in Figure 14.26. The HART protocol permits all digital communication with field devices in either point-to-point or multidrop network configuration. Figure 14.27 highlights the optional *burst* communication mode where a single slave device can continuously broadcast a standard HART reply message. Higher update rates are possible with this optional digital communication mode and use is normally restricted to point-to-point topologies.



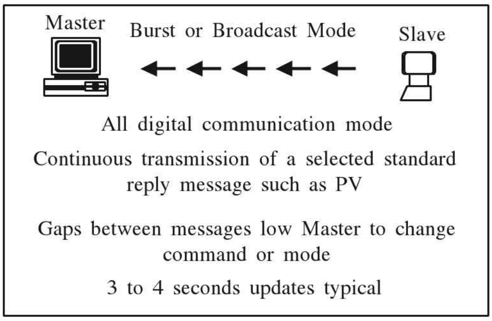
Analog + Digital or Digital only Communication

Analog signal *is not* interrupted

Slave responds to commands/requests from Master

Typical 600 ms responses (2 values per second)

Figure 14.26 Normal mode communication.



All digital communication mode

Continuous transmission of a selected standard reply message such as PV

Gaps between messages low Master to change command or mode

3 to 4 seconds updates typical

Figure 14.27 Burst communication mode.

Considerable installation savings are possible with the multidrop networking capability of HART which allows multiple field devices to be connected to the same pair of wires. In multidrop applications, communication with field devices is restricted to digital only as the loop current is fixed at a minimum value and loses any meaning relative to the process. From an installation perspective, HART signals are carried over the same wiring as is typically used for conventional 4–20 mA analog instruments today. As one might expect, cable run lengths can be longer for individually shielded twisted pairs, than those for multiple twisted pairs with an overall shield as reflected in Table 14.8. For short distances, unshielded cables may be used. HART compatible intrinsic safety barriers and isolators are also available which pass the HART signals for use in hazardous areas.

Table 14.8 Cable specifications

Distance (m)	*Cable type*	*AWG* (mm)
≤ 1524(5000 ft)	Multiple twisted pair with common shield	24(0.2)
> 1524 > 3048(10,000 ft)	Single twisted pair with shield	20(0.5)

The HART Command Set is organized into three groups and provides read/write access to the wealth of additional information available in smart field instruments employing this techology. Universal Commands must be implemented by all HART devices and provide interoperability across the large and growing base of products from different suppliers supporting the HART technology. Universal Commands provide access to information that is useful in normal plant operation such as the instrument manufacturer, model, tag, serial number, descriptor, range limits, and process variables. Common Practice Commands provide access to functions which can be carried out by many devices though not all, and Device Specific Commands provide access to functions which may be unique to a particular device. Figure 14.28 highlights the type of information that can be obtained from these devices. The integrity of HART communication is very secure as status information is included with every reply message and extensive error checking occurs with each transaction. Up to four process variables can be communicated in one HART message and each device may have up to 256 variables.

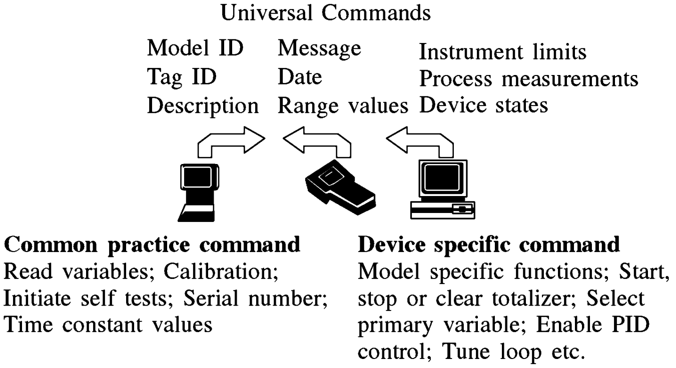

Universal Commands
Model ID
Tag ID
Description
Message
Date
Range values
Instrument limits
Process measurements
Device states
Common practice command
Read variables; Calibration; Initiate self tests; Serial number; Time constant values
Device specific command
Model specific functions; Start, stop or clear totalizer; Select primary variable; Enable PID control; Tune loop etc.

Figure 14.28 HART commands.

Device Description Language (DDL), a recent enhancement to the HART technology, extends interoperability to a higher level than that provided through the Universal and common Practice Commands. DDL provides a field device (slave) product developer with the means to create a complete description of their instrument and all relevant characteristics, such that it can talk to any host device using the language. This is analogous to a printer driver in the personal computer world which enables an application to talk with a printer such that what gets printed on the page is what was expected by the application. Universal hand-held communicators capable of configuring any HART-based instrument through DDL are available. Broader application in other types of host systems is expected. The HART Communication Foundation manages the centralized library of all registered Device Descriptions and DDL is being supported by all members of the Foundation.

The benefits of HART

The relative simplicity of the HART protocol makes it easy for both end-users and suppliers to gain experience and benefit from the enhanced two-way communication capability of smart field instruments using this technology. Powerful multiparameter instruments, efficiency with remote communication, field device diagnostics, cost effective control in field devices, installation savings with multidrop networking, and flexible/accurate digital data transmission are all achievable with instruments that use HART Field Communications Protocol. And, the compatibility with 4–20 mA makes it easy for users to incrementally add HART speaking instruments, incrementally protecting their investment in existing systems and gaining the benefits of enhanced field communication.

The advantages of enhanced field communication enabled by HART technology have been quantified in terms of maintenance, installation, and commissioning cost savings in a wide variety of applications in industries. Some of the many reasons cited for the tremendous growth and acceptance of HART technology include:

- The wide variety of products available today from major instrumentation suppliers around the world. Users have the freedom to choose the right product for their application and interoperability is assured by the common command and data structure.
- Relatively easy to understand and use, the HART protocol provides access to the wealth of additional information (variables, diagnostics, calibration, etc.) available in smart field devices employing this technology. HART enables field instrument suppliers to incorporate powerful features into their products such as PID control algorithms, diagnostics, and additional process measurements. User access to these features is provided through the enhanced communication capability of HART.
- HART is a no risk solution for enhanced field communication. For maintenance and operations people, HART is a relatively easy transition especially in point-to-point applications. Any fears about being able to keep the plant running are small as the 4–20 mA signal can still carry the process variable (as with traditional instruments), and the enhanced two-way field communication capability of HART can provide real benefit for improved plant performance.

- And, in applications where appropriate, the multidrop capability of HART provides the opportunity to connect several field instruments on the same pair of wires, substantially reducing installation costs.

14.3.9 Token Buses and Rings

The MAC functions of token rings and buses are similar. ARCnet, developed by the Datapoint Corp. in the 1970s, is a token passing protocol that can be implemented in a bus or star topology using coaxial or UTP cables. A *token* is passed around the bus or ring. Whichever node has the token, it is allowed to communicate on the media. ARCnet runs at 2.5 Mbps, at the following lengths:

- 400 feet with 10 nodes using UTP
- 2000 feet with a practical limit of up to 100 nodes using RG-62 coaxial cable in a coaxial star configuration (using a hub or hubs)
- 1000 feet with 10 nodes per 1000 feet segment using RG-62 coaxial cable.

ARCnet uses active and passive hubs in the star configuration, with network cards on the devices that have switches for setting node numbers. The lowest numbered node is the master controller, giving permission to communicate to each node by number. ARCnet is available in a 20 Mbps version.

The IBM token ring protocol, standardized via IEEE-802.5, runs at 4 or 16 Mbps. Nodes on the ring connect to a Multi-station Access Unit (MAU), a type of hub. MAUs can be connected together in a main ring, with segments or lobes from each MAU connected in a star configuration to devices with network interface cards. The length of the ring is limited to 770 metres and the maximum number of nodes allowed on a ring is 260 using STP cable. STP cable (150 ohm) is used most often, but UTP (100 ohm) cable can be used if passive filtering is provided for speeds up to 16 Mbps. Bridges can be used to connect rings. A *bridge* belongs to both layer 1 and layer 2 of the OSI model (Figure 14.29). It allows the connection between two different sections of a net that have different electrical and coding characteristics.

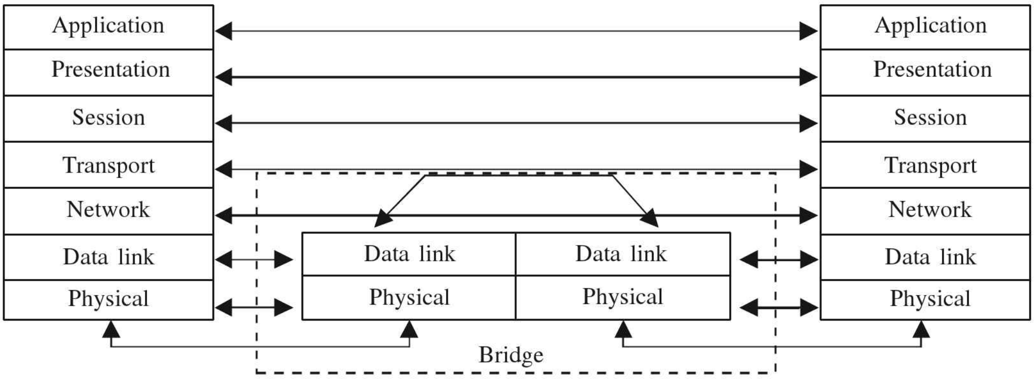
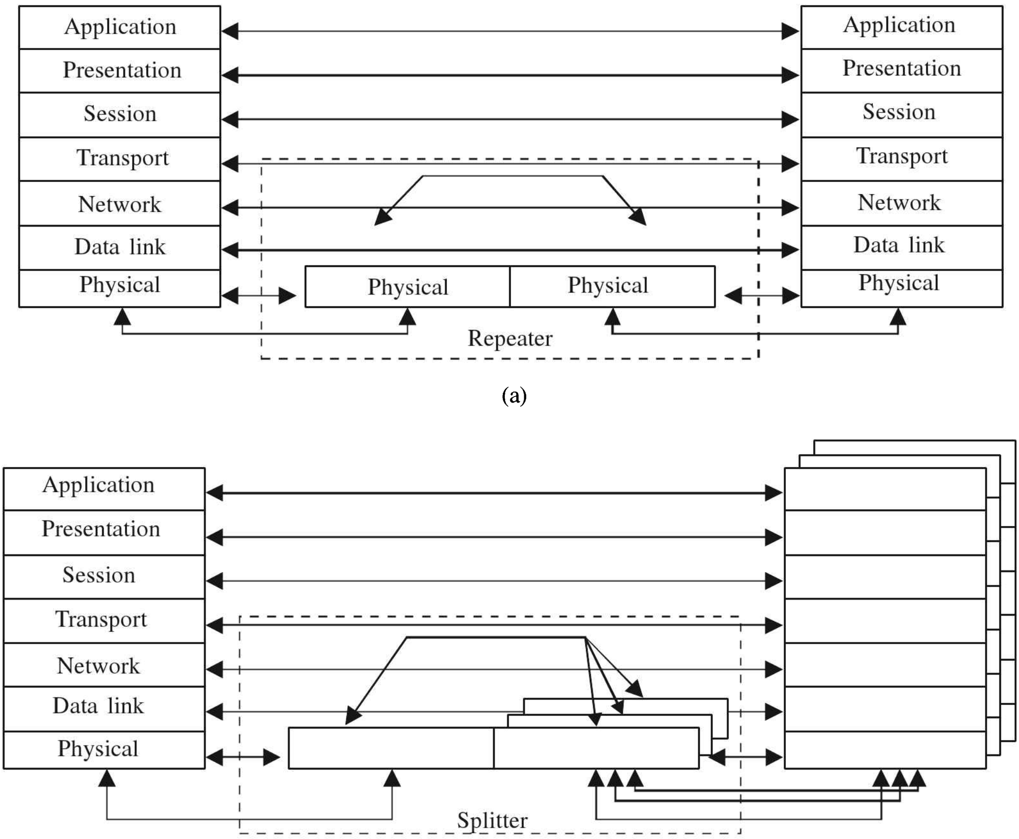


Figure 14.29 Function of a bridge.

Jitter is an interesting problem that can arise on token ring networks, where nodes that are supposed to be synchronized with the master node receive distorted waveforms due to cable attenuation. The result is that each node operates at a slightly different speed. Jitter restricts the number of nodes allowed on the ring (72 at 16 Mbps on UTP). Jitter suppressors are available that can help alleviate this problem.

Repeaters are available for extending the ring. Using the phase-locked-loop (PLL) technology, a repeater can extend the main ring an additional 800 feet at 16 Mbps on Category 5 copper UTP. A *repeater*, or amplifier, is a device to enhance electrical signals, to allow them to travel greater distances between the nodes and connect a larger number of nodes to the net. In agreement with the OSI model, this device encodes and decodes signals only at Layer 1, the physical layer (Figure 14.30).



(a)

(b)

Figure 14.30 Function of (a) repeater and (b) splitter or modified repeater.

Using a media converter or fiber optic transceiver, conversion between copper and single mode or multimode fiber is possible, thus extending main ring lengths or lobe lengths up to 2 km.

14.3.10 Ethernet or CSMA/CD

The most common bus topology in business applications, and now in industries too, is Ethernet. Ethernet originally was developed by the Xerox Corp., and later improved in association with Digital Equipment Corp. (DEC) and Intel. IEEE 802.3, based on the Ethernet standard, specifies a Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Media Access Control (MAC) protocol. This network standard provides functionality at the first two OSI network layers the physical and data link layers.

This MAC protocol allows independent transmission by all nodes on a network segment. A node sends messages down the segment with data, addressing, and control bits. All other nodes *see* the message, but only the node with the destination address will acknowledge and receive the message. When the sending node is using the segment (transmitting), all other nodes with messages to send *hear* the carrier and do not send. This type of protocol is known as *listen before send.* If two nodes attempt to send at the same time, a collision is detected, and both nodes stop sending, wait for a small, random amount of time, and then resend. Where there is a lot of network traffic, many collisions can occur and overload the system.

The Ethernet CSMA/CD standard can be implemented in five different ways:

- 10Base5—standard “thick” coaxial cable
- 10Base2—“thin” coaxial cable
- 10BaseT—UTP
- 100BaseT—UTP
- 100BaseF—fiber optics.

The XBaseY nomenclature is interpreted as follows: X = the signaling rate in Mbps, and Y = some indication of the media type. The “Base” portion of the term means that the signal uses the cable in a baseband scheme (as opposed to a broadband, multi-frequency, multi-channel modulating scheme).

10Base5 is a standard Ethernet implementation that uses RG-58 coaxial cable segments up to 500 metres long. Transceivers are clamped on to the segment, with a minimum distance between transceivers of 2.5 metres. A transceiver cable (or spur) connects the transceiver to the network node or device; this cable can be up to 50 metres in length. The maximum number of transceivers allowed is 100 per segment, and both ends of each segment are terminated with a 50-ohm resistor. Transceivers can be connected to hubs for fanning out to network nodes.

The 10Base2 “ThinNet” does not use a long coaxial segment with transceivers. Instead, modular repeaters are used with ThinNet modules in a star configuration. Thin coaxial cable is terminated with “T” connectors on network nodes, with multiple nodes connected in series on a segment, with one end of the cable home-run back to the repeater. The maximum length of a series segment is 185 metres, with a minimum distance between nodes (T-connectors) of 0.5 metres. No more than 30 nodes are allowed on a segment. 50-ohm terminators are required at each end of a segment.

The 10BaseT implementation uses UTP cable, with best results from Category 5 copper. Nodes are connected to a hub in a star configuration. The maximum length of one segment is 100 metres.

The 10BaseF implementation uses 50, 62.5, or 100-micron multimode fiber optic cable. Fiber transceivers are available for connecting to network interface cards that do not support fiber.

Repeaters can be used to connect 10Base5 segments together up to 2500 metres, and components and smaller networks can be linked together with interface devices to join the various media types into a larger network.

The IEEE-802.3 standard covers several CSMA/CD implementations, with lengths up to 925, 2500, and 3600 metres. The number of nodes allowed follows the hardware and performance requirements of the network operating system. More nodes and higher traffic mean more potential for collisions. Network performance might drop when you need it the most, such as during a plant transient when data flow and control element actuations are high.

Ethernet or CSMA/CD is established for 1, 10, 100 Mbps and now 1Gbps. 100BaseT is essentially 10BaseT, only 10 times faster. They both use the same MAC layer. At higher speeds, implementations account for round-trip delays for collisions. Table 14.9 gives vital statistics of Ethernet at different speeds.

Table 14.9 Vital statistics of Ethernet

Parameter	*Ethernet*	*Fast Ethernet*	*Gigabit Ethernet*
Data rate	10 Mbps	100 Mbps	1000 Mbps
CAT5 UTP cable length	100 m	100 m	25–100 m
STP/coaxial cable length	500 m	100 m	25–100 m
Multimode fiber cable length	2 km	412 m (Half duplex) 2 km (Full duplex)	500 m
Single mode fiber cable length	25 km	20 km	2 km
IEEE Standard	802.3	802.3u	802.3z

Finally, a wireless Ethernet implementation is available. Line of site is not required indoors. Wireless repeaters and bridges are available to link cells together.

Decisions regarding local area network (LAN) topologies and MAC protocols can affect the way a network is managed for years to come. A new facility can be wired with state-of-the-art network components and premise wiring devices, but the network administrator or plant engineer is usually faced with an existing infrastructure, where options are constrained by installed media. For example, an old, abandoned 4–20 mA dc current loop STP cable can always be used to implement 10BaseT Ethernet and thus save considerable project funds. Or, a new fiber optic cable can be purchased with multimode fibers for a low bandwidth project and still have spare, single-mode fibers for future, high bandwidth applications.

Token passing and CSMA/CD protocols make interesting comparisons. Token passing offers a predictable access window with consistent performance on large, busy networks, but nodes need to wait for the token, and large rings can result in significant delays. CSMA/CD allows a node to transmit immediately when the network is quiet, but performance can be unpredictable and depends on the probability of collisions.

Upgradeability and growth are two areas that should be considered for each new segment or modification to a given segment. The challenge is to keep an eye on network demands as new applications come along, and select the right amount of bandwidth and upgradeability.

14.4 MOVING UP THE LAYERS

So far, the physical layer has been described with implied functions at the data link and network layers. Two devices can communicate by way of a simple RS 485 link or they can be part of a larger LAN with complex media access controls.

The data link's only function is to pass data from one node to the next. The data link layer can connect two nodes or devices with a *bridge* (Figure 14.29). Early bridges only connected two segments of a network with the same protocol (such as CSMA/CD or token passing). Smart bridges are available now that can connect differing protocols, with selective forwarding of data packets. Various levels of data link performance exist depending on the size of the network (wide vs. local area network; WAN versus LAN) and the protocol/media selections.

The network layer routes incoming data for another node onto an appropriate outbound path. The physical device that performs this function is called a *router*. Since bits still need to be pushed down some physical media, a router inherently performs the physical and data link functions in order to go up and down the OSI model (Figure 14.31), although routers are not sensitive to the data link and physical layer details, allowing them to connect different types of networks. A router takes an incoming packet, looks at the destination address, determines the best route, and provides the required addressing. One common implementation of this function is Transmission Control Protocol/Internet Protocol (TCP/IP). On Novell networks, this function is called Sequenced Packet Exchange/Internetwork Packet Exchange (SPX/IPX).

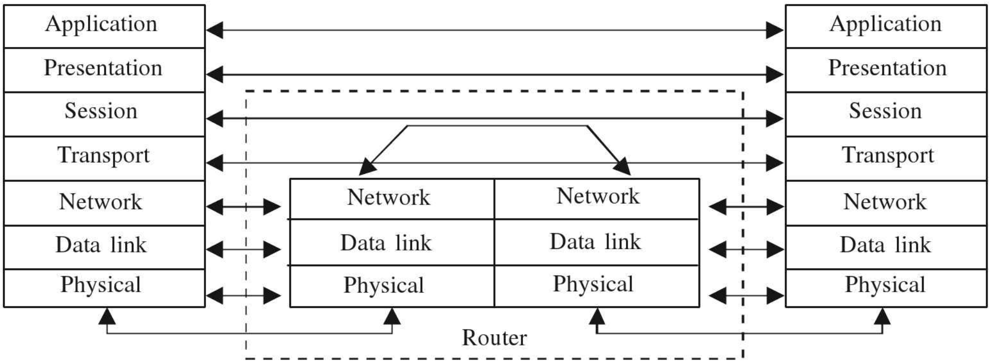


Figure 14.31 Function of a router.

The TCP and SPX protocols operate at the transport layer, and the IP and IPX layers operate at the network layer. Thus, routers send and receive packets between nodes with "IP addresses." The downside is that routers can only exchange data between nodes using the same network layer protocol.

Gateways operate at the network layer or above; they are used to connect dissimilar networks or systems via network protocol conversion. The term *stack* refers to pieces of the OSI model that are bundled together. In client-server models, the TCP/IP software is sometimes referred to as a stack.

For example, on a PC with a network interface card, a word processor program (application), operating system (Windows NT), and redirector are at work (redirectors intercept network application I/O requests alongside a device's operating system). In this example (Figure 14.31), a user wants to open a file stored on a remote file server on another network by clicking on the G drive in the list of available drives (including the local A, B, and C drives). The redirector sees the I/O request, recognizes its external location, and sends it to the network interface card. On its way, the request goes through the TCP/IP stack where addressing and protocols at the transport and network layers are handled. The network interface card receives the request and holds it until it can get on the network and pass it on (push the bits) to the next node. The request is sent from node to node, using appropriate IP addresses and protocols until it gets to the server, where it is processed, and the reply is sent back essentially the same way.

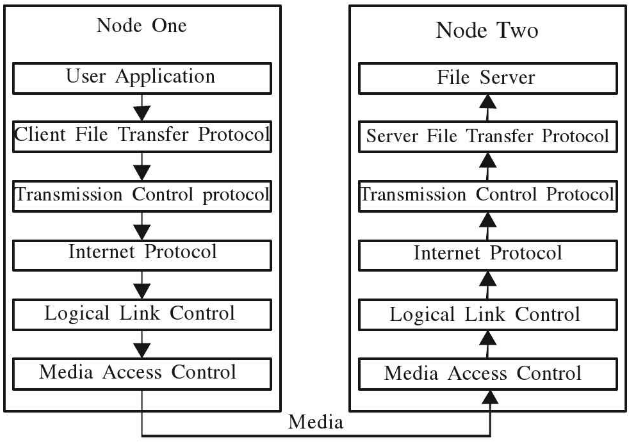


Figure 14.32 A remote file request according to OSI.

Backbones

In larger enterprises, some sort of information backbone or data highway allows high speed and high quality internetworking with good reliability. Backbones connect LANs, WANs and other forms of network segments together in one large network. The decision to build a backbone depends on the present and future needs of the enterprise. But once it is implemented, growth becomes less painful. One common backbone is the Fiber Distributed Data Interface, or FDDI standard. This backbone is also available in a copper version (CDDI).

FDDI is considered an implementation of a *metropolitan* LAN, or MAN. It is a dual-ring topology that can span up to 200 km at 100 Mbps. It is commonly used to connect LANs together where higher bandwidths are required. Two rings of fiber carry data. All nodes attach to at least the primary ring, with some or all attached to the secondary ring. The idea is that if one ring breaks, the other automatically picks up the load. If both rings break at the same point, they automatically join together in one long ring (Figure 14.33).

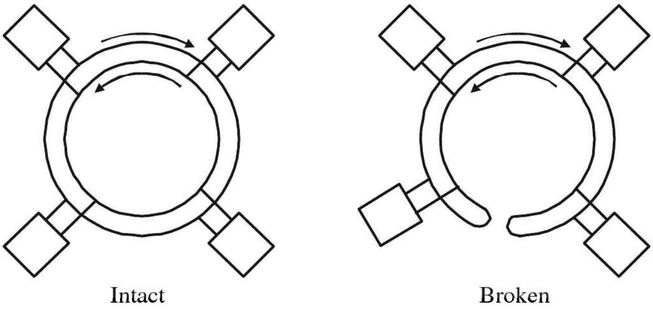


Figure 14.33 Dual ring redundant FDDI network.

FDDI hubs are available that bring together various FDDI devices, LANs, and Ethernet devices over single-mode or multimode fiber or copper. FDDI is a good choice for linking buildings together on a campus or industrial site.

Another implementation of an enterprise backbone, either stand-alone for one particular link or in conjunction with other technologies, is microwave technology. Operating at frequencies between 2 and 40 GHz, transmitters and receivers use long, line-of-site transmission paths. The economics are reasonable when compared with trenching and laying cable. The higher the tower, the longer the link (of the order of tens of miles). Data rates are typically 1.544 Mbps. Leasing telephone lines is usually the alternative to microwave.

Satellite technology is also used for industrial and commercial network implementations. Remote transmitters for stationary platforms (such as well heads) or mobile platforms (as in the trucking industry) are being used to transmit real-time data via uplinks to a communications satellite and back down to ground stations. Telephone companies provide switched services between the satellite system and a given enterprise network. A typical Enterprise Network is shown in Figure 14.34.

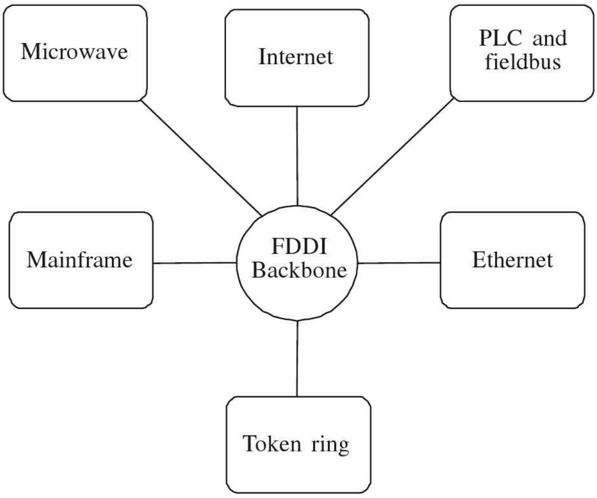


Figure 14.34 A typical enterprise network.

14.5 FIELDBUS AND DEVICE NETWORKS

While the network technologies described so far are typical for business information systems, they are converging with industrial instrumentation and control systems. Industrial plants have used information technology (IT) for years, but with open IT standards, faster computers, emerging software, and demand for integrated information by all segments of an industrial enterprise (accounting, engineering, operations, maintenance, and management), industrial networks are merging right into enterprise-wide IT solutions.

PLCs and distributed control systems (DCSs) are offering Ethernet/TCP/IP connectivity so that real-time information on plant processes is readily accessible by any workstation on the network (LAN or WAN), in a client-server relationship. A plant manager can watch a graphic display of plant operations in one window while scanning accounting data in another. Object Linking and Embedding (OLE) and Dynamic Data Exchange (DDE) can merge data streams, such as unit cost (from accounting) and production totals (from the plant floor) in one spreadsheet. Internet and intranet browsers running Java or HTML applications access data acquisition and control systems using standard LAN/WAN technologies.

Meanwhile, on plant and factory floors, industry has developed its own range of fieldbus or device-level networks for linking control devices with increasingly intelligent instrumentation. As the terms have evolved over the past several years, device networks typically embrace those optimized for discrete manufacturing automation, while fieldbus connotes more process-oriented instrument networks. On the discrete side, the emphasis is on high-speed transmission of typically smaller information packets. Process fieldbuses, on the other hand, sacrifice speed for more secure transmission of larger information packets. Intrinsically safe capabilities also are prerequisite for many process applications.

14.5.1 The CIM Pyramid

Before choosing the solution for an industrial problem, it is necessary to perform an abstract analysis, independently of any economical, political and commercial consideration, no matter how difficult it is to avoid the latter.

In a factory or manufacturing plant there are several levels of data processing, for each of which requirements differ. Administration has megabytes of data, but response time is not critical and lie in an interval from a few minutes to hours or even days (backups are performed once a day, written reports to managers can tolerate a few hours delay with no harm.)

Production, instead, which is the main user of field buses, has different constraints. At the lowest level, we find buses that transfer just a few bytes, but the reaction time must be of the order of a few tens of microseconds.

Such differences define a first criteria to evaluate the choice of the field bus type to adopt. Parameters such as response time and quantity of data are generally known for a given application, even though the latter cannot easily be integrated in a well structured concept such as the Computer Integrated Manufacturing (CIM) pyramid (Figure 14.35).

Observe that this representation is ideal and has to be used only as a reference. The reality of companies does correspond little to the integration model, because the critical mass to justify such investments is rarely reached.

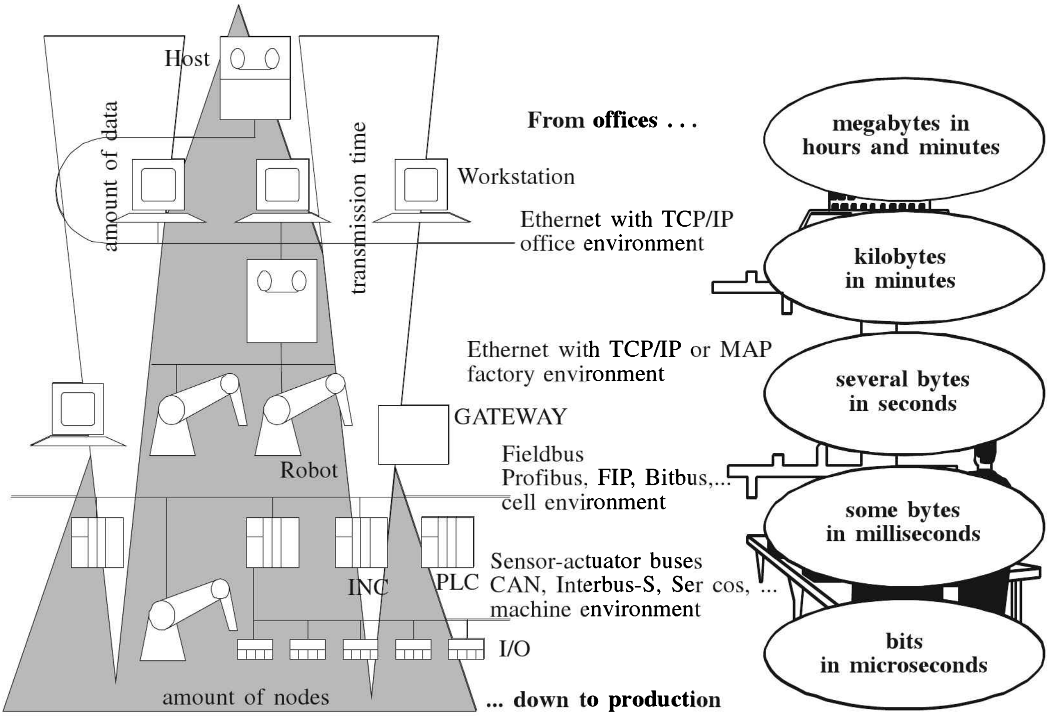


Figure 14.35 The CIM pyramid.

14.5.2 What is a Fieldbus?

Fieldbus is a generic name given to fully-digital communication protocols for industrial measurement and control applications. Ideally, the cables used are more closely related to communication technology than to process instrumentation, but operation is also possible over conventional process field cabling, at least over limited distances (say up to 400 m). The intention is that the same protocol should be usable, with appropriate communication speeds, for both process control and factory automation (machinery control). The IEC and the ISA (now called the International Society for Measurement and Control) have more or less coordinated their efforts to reach a worldwide international standard, which is IEC 1158 and ISA S50.02. For process control, the significant features of the IEC/ISA fieldbus are:

- Higher communication speeds (31.25 kbps, 1 Mbps, and 2.5 Mbps)
- Operation over two wires, also carrying power to field devices
- Multidrop operation (up to 30 devices on one pair of wires)
- Possible operation in hazardous areas using intrinsic safety techniques
- Interoperability between equipment from different suppliers.

Note that HART offers all these features except the first. HART (Highway Addressable Remote Transducer) is also digital communication protocol which operates on top of a conventional 4–20 mA current loop signal from a measurement device—a transmitter—or to a valve positioner. HART uses a low-level frequency-shift-keyed sine-wave signal to convey the digital communication. This has an average value of zero, so does not affect the analog signal. Because the analog current loop is typically unbalanced (grounded on one side), and not impedance-matched, cable capacitance attenuates high-frequency signals. HART therefore uses relatively low frequencies (1200 and 2200 Hz) and this, in turn, restricts it to 1200 baud (equivalent to 1200 bits per second, with simple "1-bit-per-symbol" coding techniques). Similar modulation methods used to carry digital signals over telephone lines have somewhat more capacity, because they do not have to contend with the possible levels of cross-talk which can occur in the multi-pair cables commonly used in industrial process field wiring.

HART has typically been used for remote configuration, adjustment and diagnostics of smart field devices, and for this it is very adequate. Generally where fast response is important, it is not a good idea to use HART communication to read measurement values; the continued existence of the analog signal is important in such cases. So is the continued compatibility with existing analog connections on most control systems (SCADA and DCS). But in some applications, where high data rate is not required, HART has also been successful in its multidrop mode (up to 15 devices). In this mode, the analog signal is not used, and all measurements are made using HART digital communication. Even when using the analog signal for the measurement, we can gain a valuable extra degree of confidence in that measurement by continually checking the status of the field device by HART communication. HART-compatible DCSs usually do this.

Field devices are becoming more complex, *smarter*, and hold more information about themselves and the process to which they are connected. A higher communication data rate is important in handling this additional information without undue delays. Fieldbus (at 31.25 kbps) provides this capability, and is also useful for measurement of process pressures, temperatures and flow rates. The higher-speed variants (at 1Mbps and 2.5 Mbps) are necessary for machinery control and in PLC applications where millisecond response times are required.

14.5.3 Why Define a Fieldbus?

Fieldbus is characterized by the following criteria:

- Completely digital replacement of 4–20 mA
- Control, alarm, trend and other functions distributed to devices in the field
- Interoperable multivendor
- Open system; specification available without licensing agreement

Digital communication

Fieldbus is a complete system, with the control function distributed to equipment in the field, while still allowing operation and tuning from the control room using the digital communication. It replaces the traditional 4–20 mA and the classic DCS where the control function was centralized to one or more 'control cards' used in the central computer as shown in Figure 14.36. But this does not meet the four criteria above.

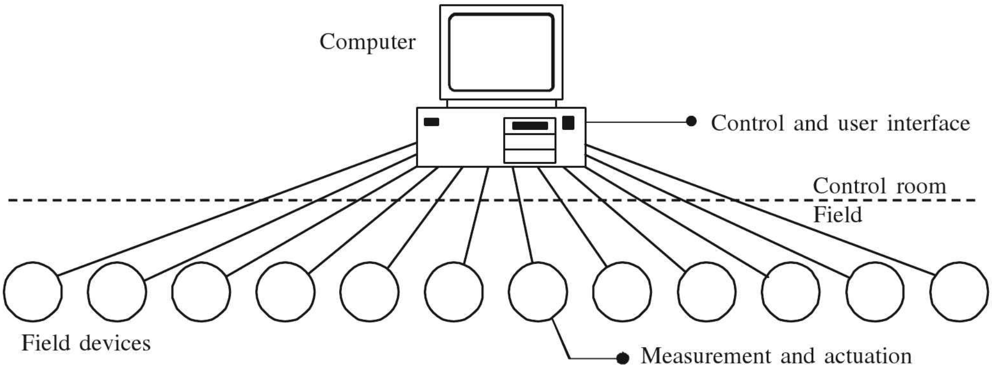


Figure 14.36 Direct Digital Control (DDC).

Figure 14.37 shows the Distributed Control System (DCS) where the control is partially distributed to a few control cards, still in the control room, each having several loops.

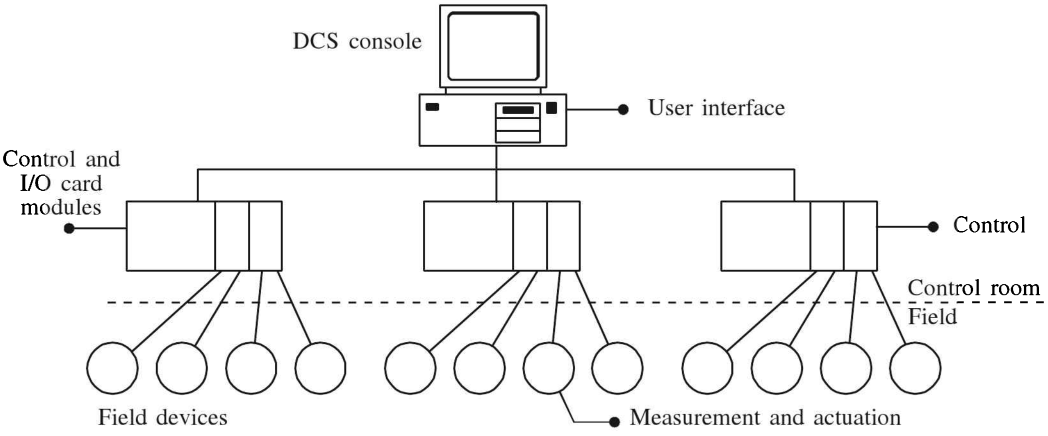


Figure 14.36 Distributed Control System (DCS).

Figure 14.38 shows the Fieldbus system where control is totally distributed to field with loops in individual devices.

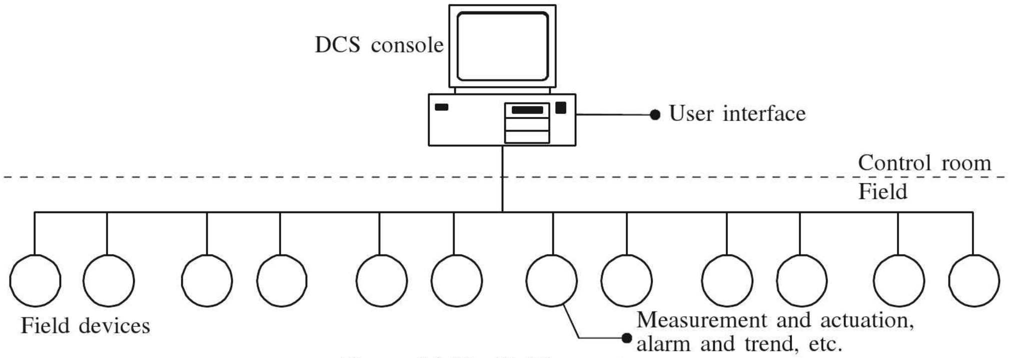


Figure 14.38 Fieldbus system.

Some advantages of bi-directional digital communications, over 4–20 mA, and other good features are known from existing smart transmitter protocols:

- Higher accuracy and data reliability
- Multi-variable access
- Remote configuration and diagnostics
- Reduction of wiring
- Use of existing 'analog wiring'

Accuracy. Higher accuracy can be attributed to digital communication, since the microprocessors in, for example, a transmitter and a controller may talk directly, rather than going through D/A and A/D conversions, of which there may be many in a closed loop. Status is passed along with measurement and control data. It is therefore possible to determine if the information is reliable or not. All data are checked and guaranteed free from distortion due to noise or an impedance mismatch that may affect analog undetected signals.

Multivariable access. Multivariable access, meaning, for example that a pressure transmitter is not limited to a single output for pressure, but also informs process temperature. Another example is access to setpoint and manipulated variable of a controller in the same device, or multiple channel inputs for a temperature transmitter. The multivariable capability of Fieldbus allows control, totalization and other signal processing in the field. Therefore, a separate controller or other signal conditioning equipment is not necessary. The host may be a simple off-the-shelf PC with MMI software.

The multivariable access virtually floods the control room with information. Classic recorders may not be able to handle the job. Paperless recorders and computer storage may take over. Such information may be used for statistical process control and other process management.

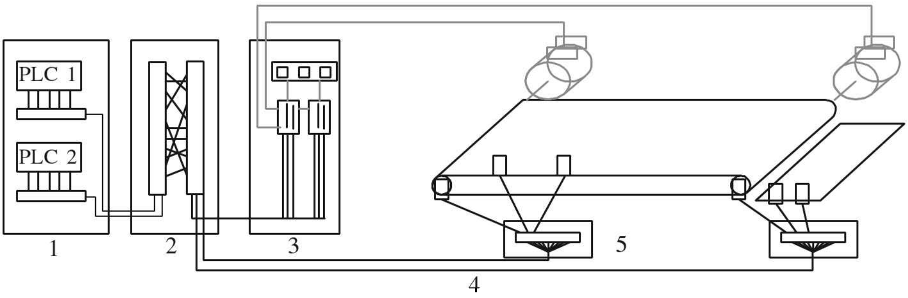
Fieldbus has, on top of this, software function blocks which replace many functions today performed by hardware. This provides tremendous flexibility since the control strategy may be edited without having to rewire or change any hardware. Once physically connected, logical connections between function blocks may be changed, function blocks can be added and removed. More advanced devices may execute a virtually unlimited number of function blocks. If a system has to be expanded or improved, the need for additional hardware is minimized, just by letting existing devices execute more blocks.

Remote configuration and diagnostics. The digital communication allows the complete configuration to be changed remotely. Calibration can be done in operation without having to apply any input or measure the output. Similarly, the status of the self-diagnostics may be interrogated.

Wiring. Reduction in wiring and simplification is achieved through connection of several devices on a single pair of wires—multidropping. Connection is a simple task, since everything is in parallel and terminal number matching is at a minimum. This means low cost and easy replacement of old transmitters. Multidropping of several devices on a single wire may drastically reduce the amount of needed cable. In many factories, a device can be a kilometre or even further away from the control room. Since every loop needs at least two pairs of wires (one for the transmitter, the other for the actuator) a refinery can have several hundreds of such loops. In all, the cabling saving for a medium or large factory is immense. Transmitters and actuators are often located next to each other, but far from the operator console, an ideal situation for multidropping. The reduction in the number of devices and wiring, with associated cable trays and marshaling boxes yields a less costly system.

Another example is shown in Figure 14.39. In this example, two programmable logic controllers, PLC's, are controlling a process. The process consists of the conveyor and other equipment. For connecting the PLC with the traditional method [Figure 14.39a] we need (1) PLC cabinet, (2) cross connection cabinet, (3) cabinet for variable speed drives, (4) a number of shielded multiconductor cables, and (5) process cabinets for terminating the cables. Using the fieldbus, we need not use the cross connection cabinet, variable speed drive cabinet and process cabinets. For the installation, we will need only the PLC cabinet and the fieldbus cable [see Figure 14.39b].

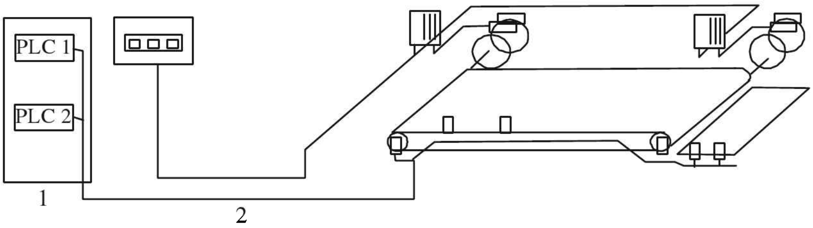
A traditional wiring from the PLC to the process



1. PLC-cabinet
2. Cross connection cabinet
3. Variable speed drive cabinet
4. Shielded multiconductor cables
5. Process cabinets

(a)

A fieldbus connecting the PLC and the process



1. PLC-cabinet
2. Fieldbus cable

(b)

Figure 14.39 Connecting PLCs with the field devices.

The slowest option for Fieldbus is 25 times faster, and far more efficient, than the most common smart transmitter protocol, ensuring tight closed-loop control. The low speed version of Fieldbus was designed to use the same type of wiring as analog and smart transmitters, being able to easily replace them. However, it should be noted that in order to use the multidrop feature, transmitters must be connected in parallel. Figure 14.40 to Figure 14.42 elaborate the difference in connections between controllers and transmitters and actuators using analog, analog and digital and only digital signals as in Fieldbus. In Figure 14.40 each individual variable has a physical 4–20 mA connection, in Figure 14.41 individual variables involved in closed-loop control have a physical 4–20 mA connection. Superimposed on it is a digital signal for configuration and diagnostics whereas in Figure 14.42 a single physical wire is the medium for multiple variables having logical connections.

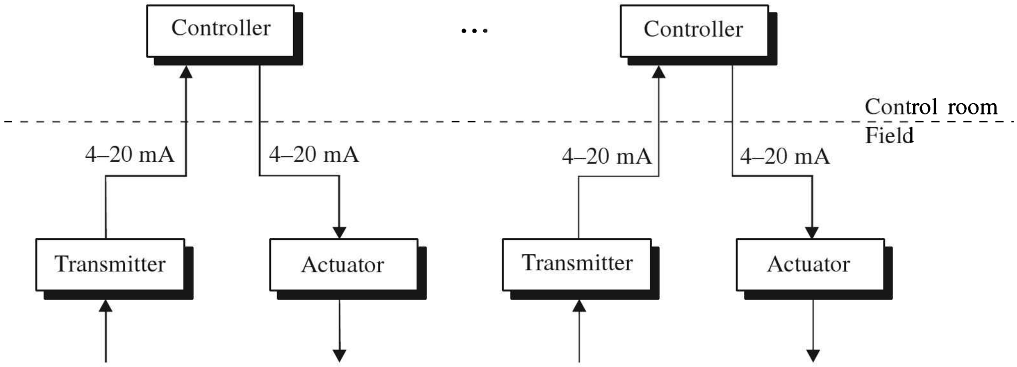


Figure 14.40 Connections with analog signals.

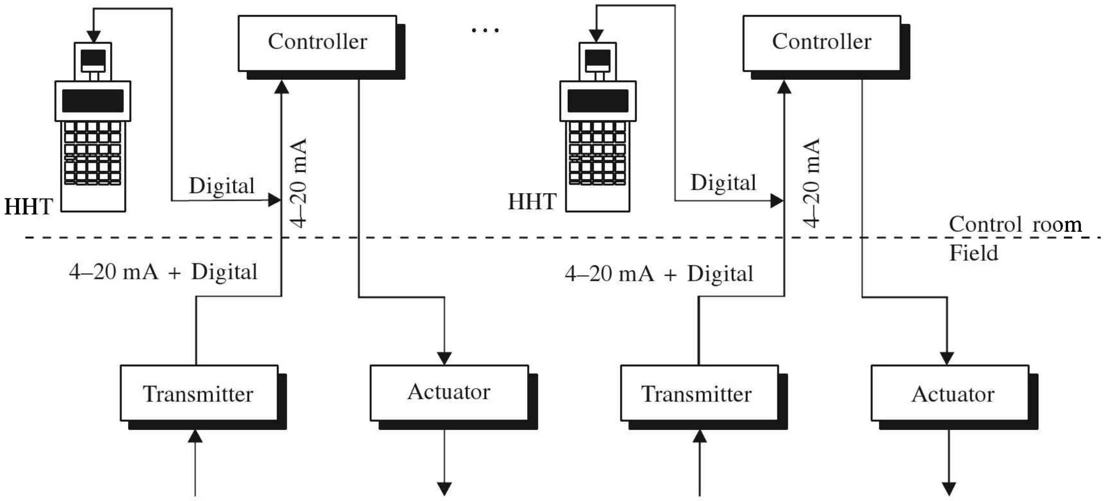


Figure 14.41 Connections with intelligent/smart hybrid of analog and digital signals.

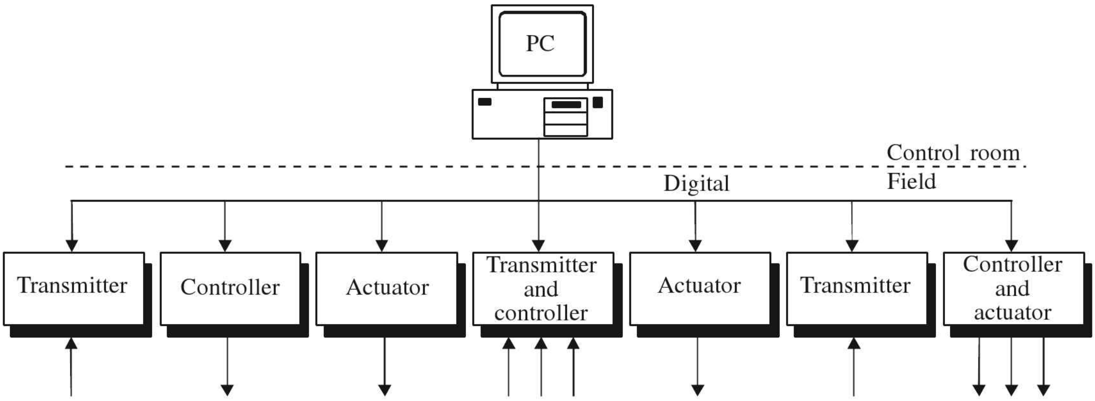


Figure 14.42 Connections with Fieldbus pure digital communication.

Open system and interoperability

Many smart transmitter protocols are proprietary and unique to a single manufacturer. This lack of standard means that the user is locked to a single manufacturer, if he selects their system. If that manufacturer cannot provide an urgent replacement, or does not have a particular type of transmitter, the user's only option is to return to 4–20 mA. The user depends on a single vendor that cannot provide the latest technology and features. The vendor is in a position where he can set any price.

The ideal opposite to a proprietary system is an *open* system. Open systems are based on off-the-shelf standards enabling multiple vendors to provide interoperable hardware and software.

The ability of 4–20 mA devices to replace any other device of the same type is called *interoperability*, roughly meaning compatible. Fieldbus offers the same capability. A brand 'Y' transmitter can be replaced by a brand 'X' transmitter of the same type any time, without loss of functionality, and can interface to another brand 'Z' device. Fieldbus forces interoperability between the devices, which are complying with this standard and it is also available to all manufacturers and users without licencing agreements. It is *open*, it is fully disclosed, there are no *secrets*. The user or a third party can make their own configurations and software. Users may now select a device based on price, performance, quality, and delivery time. They may mix and match the best of each type, just as they could with the 4–20 mA. They do not have to choose a device manufactured by a certain company just to match other devices of the same brand already installed (without Fieldbus, users would have been forced to develop special communication drivers in this case). Another benefit of interoperability is that system software does not have to be upgraded when new products are introduced.

The lack of standard means that smart features of existing intelligent field devices are largely unused. Communication is only used for calibration. For example, there are very few existing systems that have managed to make use of the multidrop capability of existing smart transmitters. Though smart devices have self-diagnostics, status is only informed when device is polled. This is rare in most applications since communication is normally done only at calibration. One must suspect and query for a failure before one finds out about it.

The impact on the user, when turning to Fieldbus is great. The main points are:

- Even higher reliability operation
- Virtually unlimited flexibility
- Reduction in equipment cost
- Reduction in installation cost
- Mass of information

14.5.4 Fieldbus Requirements

The main requirements for Fieldbus are to overcome the problems of smart transmitter protocols while maintaining the advantages of 4–20 mA standards (the main advantage of 4–20 mA is tight closed-loop control).

By providing various options for communication speed and device powering, the requirements for both intrinsic safety and minimum communication delay can be met. By optimization of network use, tight closed-loop control can also be achieved where intrinsic safety is required.

With 4–20 mA technology, it is possible to build a control loop containing only a transmitter, a controller and an actuator. Fieldbus devices are also capable of doing so, as well as acting in a larger control system. Fieldbus devices are multipurpose and as versatile as 4–20 mA. Fieldbus devices therefore operate by themselves with a simple user interface in order to be economical in small systems. The cost of a host computer with dedicated software and certainly a DCS cannot be justified for a small system, even though costs are going down. There is also a logistics problem for both users and manufacturers if they are forced to keep using analog technology in small systems.

The possible complexity of a system where so many devices can be connected together (and where each device can perform the function of several conventional devices) requires a user-friendly interface. The user must be freed from manual address assignment, as seen in smart transmitter protocols, and the painstaking job of tracking bits, bytes, words and memory addressees, as it is done in PLCs.

The function block model is the choice of all Fieldbus proposals. The user can easily relate to it since the device is now represented by blocks, just like blocks in ISA and SAMA control diagrams. Physical wiring is now logical connections or 'soft wiring,' links between blocks. Though technically different, it appears very familiar, and users feel comfortable with it. Device address and parameter indexes are automatically assigned.

Standardization ensures interoperability, but if it is too rigid it may have adverse effects for the user. There must be room for manufacturer differentiation. If the standard forces compliance to every conceivable detail, the user would actually have nothing to choose from, because the devices from all manufactures would be exactly alike. If one manufacturer came up with a great idea, he would not be able to implement it, or forced to go through a process to have it included in the standard, whereby his competitors would learn about it.

Fieldbus specifies the basic functionality requirements, but allows a manufacturer to add unique features to their device. These features benefit the user, and the manufacturer may use them as marketing tools. Likewise, if newer models are different from their predecessors, devices communicating with it must know that. In short, Fieldbus must not hinder development and improvement of products. Fieldbus provides a mechanism that ensures that interoperability is maintained for the manufacturer specific features as well.

The cost of shutting down a system can be very high in terms of production loss. To be able to configure the system while in operation is therefore a requirement met by Fieldbus.

14.5.5 Types of Fieldbuses

Based on the application, a number of manufacturers introduced their own standard for the fieldbus. Fieldbuses at low level (Level 0) are called sensor-actuator buses. Their features are as follows:

- Connect sensors and actuators to the automation system
- Replace I/O wiring, cross connection cabinets and terminal boxes
- Time response is very short, from microseconds to a few milliseconds
- Amount of information to be transferred is very low.

For example, ASI (Actuator Sensor Interface) is a European standard developed by Siemens for Senor Actuator Bus. ASI specifications are given in Table 14.10.

Table 14.10 ASI specifications

1.	Topology:	Connected like a tree, the bus has one master
2.	Physical media	Two wire unshielded cable, same cable is used for data communication and power supply
3.	Coding and data:	Manchester coding; 5 bit address, 4 bit data protection and a parity bit
4.	Protocol:	Master writes 4 bit output and reads 4 bit input by polling each slave during a cycle. Devices from different vendors can be connected in the same system
5.	Maximum number of devices:	One master 31 slave modules
6.	Maximum number of I/O points:	Four inputs four outputs per slave. Number of points in one segment = 31 × 8 = 248. Many ASI segments can be connected to an automation system
7.	Cycle time:	Maximum 5 ms

Features of Fieldbuses at Level 1 are as follows:

- Transmission of frames from a few dozens of bits to hundreds and thousand of bits
- Time response from a few milliseconds to tens of milliseconds
- Access to a portion of layer immediately above the Fieldbus also included in some of the buses.
- Access to resources of lower layers also available in some Fieldbuses.

Table 14.11 tabulates the developers and applications of various Fieldbuses.

Table 14.11 Various Fieldbuses

Bus	*Developer*	*Basic application*
Bitbus	INTEL, 1980	Computer networks and digital telephony
MAP (Manufacturing Automation Protocol)	General Electric, 1980	Manufacturing plants
Modbus	AEG Modicon, 1980	Process control
WorldFIP	WorldFIP promoted by Honeywell, Balley, 1988	Process control
LON	ECHELONS, 1991	Building Automation, Electric Energy Distribution Network
PROFIBUS (Process Control Field Bus)	SIEMENS, 1994	Process control
SDS (Small Distributed System)	Honeywell, 1994	Process control
DeviceNet and ControlNet (uses CAN Bus as backbone)	Rockwell Allen Bradley, 1994,1995	Industrial control
CAN (Controller Area Network)	BOSCH, a car manufacturer, 1995	Automobiles
IEC (International Electrochemical Commission) /ISP (Interoperable Systems Project) SP50	ISA(Instrument Society of America) /Fieldbus Foundation backed by Fisher-Rosemount, Siemens and Yokogawa, 1992-1996	Automobiles, Process control
Foundation Fieldbus (Protocol Standard)	Fieldbus Foundation, as per British Standard DD236, 1997	Process control

The Fieldbus Foundation is a worldwide consortium of manufacturers and industry groups that have designed and manufactured an open fieldbus technology called Foundation. Software and hardware specifications have been written by design and marketing teams, and many products are becoming available that conform to this standard across many different vendors. This bus is being adopted by most of the manufacturers and users. The following section describes the features of this bus.

14.6 FOUNDATION FIELDBUS (FF)

14.6.1 Architecture

There are two major parts to the Fieldbus system architecture—interconnection and application. Interconnection refers to the passing of data from one device to another, may be a field device, operator console, or a configurator. This is the communication protocol part of Fieldbus. The application is the automation function the system performs. By standardizing part of the application, Fieldbus has gone further than any other communication standard, ensuring interoperability between conforming products.

The Fieldbus interconnection architecture is based on a three-layer subset of the architecture from the OSI (open systems interconnect) reference model (see Figure 14.43). Layers 3 through 6 are not used because Fieldbus (and most other LANs) has no interconnection between networks, which is the purpose of these layers. This simplification makes Fieldbus faster and easier to implement in devices with limited processor power, such as field instruments. Only layers 1, 2, and 7 (physical, data link, and application layers) are used by Fieldbus. Layers 2 and 7 are considered bundled together in a communication stack. The application is the functionality provided by the function blocks.

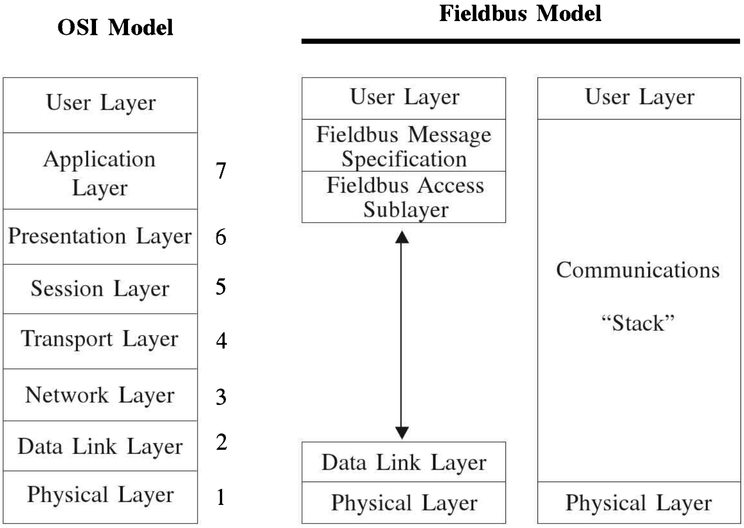


Figure 14.43 Foundation Fieldbus protocol stack.

The part of the whole application performed by the system that is performed in a device is called the application process (AP). The AP consists of two parts, one user portion, which is the functionality, and one communication portion. In Fieldbus the user portion is the actual device function, such as measurement or control (function blocks), or the user interface. A Fieldbus device has three APs—the function block application, network management, and system management.

14.6.2 Physical Layer

The Physical Layer covers the media for transmitting/receiving data from the fieldbus physical medium in terms of communications rates, signal encoding, length of connections, number of units on the bus, power supply on the bus, etc.

Closed-loop control with performance like a 4–20 mA system, and variable speed drives require high data transmission speed. Since higher speed means higher power consumption, it then clashes with the need for intrinsic safety in some applications. Therefore one moderately high communication speed was selected, and other faster non-intrinsically safe options were also made available, catering to all applications. The system was designed to have a minimum of communication overhead to meet control requirements even with the low speed option.

There are several combinations for the physical layer, each with their relative merits. All devices on a bus must use the same options for media, connection and transmission rate. However, bus or non-bus powered devices may be mixed, as well as intrinsically safe or nonintrinsically safe ones.

Wire media characteristics

The wire media uses electrical signals on a normal copper twisted pair wire, and is already approved as an IEC/ISA standard since 1992. Both fiber optics and radio communications have now been added to the IEC standard. The length of the wired bus depends on the quality of the cabling. Typical screened and twisted-pair cabling is suitable up to 1900 metres, at the low-speed communications rate of 31.25 kbps, but low-quality cabling can reduce this significantly. Similarly, use of the higher speeds reduces the acceptable wiring length. Short wiring lengths to connect devices onto the bus can be ignored, but spurs (connections > 1 metre) must be included and have rules relating to how long they may be, depending on the number of devices on the bus (Table 14.12). The fieldbus requires a terminator at each end. This is a resistance capacitance component to balance the communications network and avoid distorted communications from echoes. The communication rates are 31.25 kbps, 1 Mbps and 2.5 Mbps. The 31.25 kbps rate supports intrinsic safety (IS) applications and, as this is the priority rate for FF and suppliers, it is the initial platform for products.

Table 14.12 Maximum cable and spur lengths of a Foundation Fieldbus

Maximum cable lengths

	31.25 kbps	1 MHz	2.5 MHz	
Type A	1900 m	750 m	500 m	Single twisted pair with overall shield
Type B	1200 m			Multiple twisted pair with overall shield
Type C	400 m			Single or multiple twisted pair without any shield
Type D	200 m			Multiple conductor cable without twisted pairs

Note: Maximum lengths are test conformance lengths. Longer lengths are possible with better quality cable. Maximum lengths are fieldbus trunk + spurs.

Spurs = connections to fieldbus trunk > 1 metre.

Splices = connections to fieldbus trunk < 1 metre.

Maximum spur lengths

Devices on bus	31.5 kbps *spur length*
1–12	120 m
12–14	90 m
15–18	60 m
19–24	30 m
25–32	0 m (none)

Note: These are for a single device on the spur. The spur length must be reduced by 30 m for each additional device on the spur.

The device must isolate the communication hardware, *Media Attachment Unit (MAU)*, from ground to avoid electrical ground loops when devices are multidropped.

Bus topology (Figure 14.44), tree topology (Figure 14.45) and point-to-point topologies are supported. Tree topology is only supported by the low speed version. The bus has a *trunk* cable with two *terminators*. The devices are connected to the trunk via *spurs*. The spurs may be integrated in the device giving zero spur length. A spur may connect more than one device, depending on the length. Active *couplers* may be used to extend spur length. Active *repeaters* may be used to extend the trunk length.

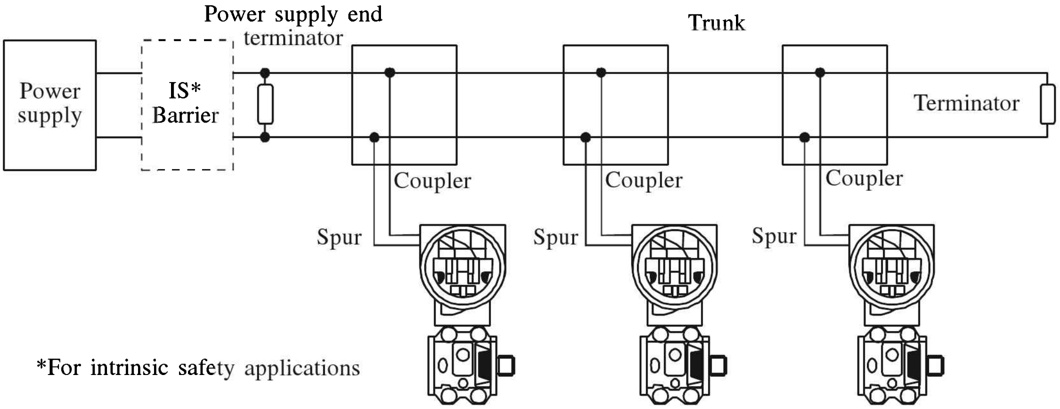


Figure 14.44 Bus topology.

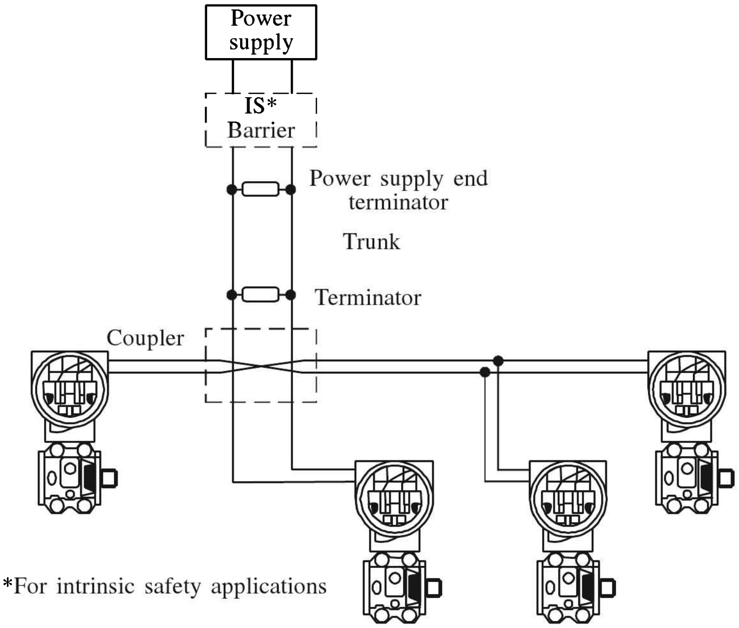

Power supply
IS* Barrier
Power supply end terminator
Trunk
Terminator
Coupler
*For intrinsic safety applications

Figure 14.45 Tree topology.

The terminators are designed to have an impedance of 100 Ω each around the transmission frequency. A device transmits by modulating current in the network according to the Manchester encoded signal. Receiving devices sense the voltage drop generated over the two terminators as the current is modulated. The modulated current is 15 to 20 mA peak-to-peak for the low speed version, with a receiver sensitivity of 150 mV (see Figure 14.46).

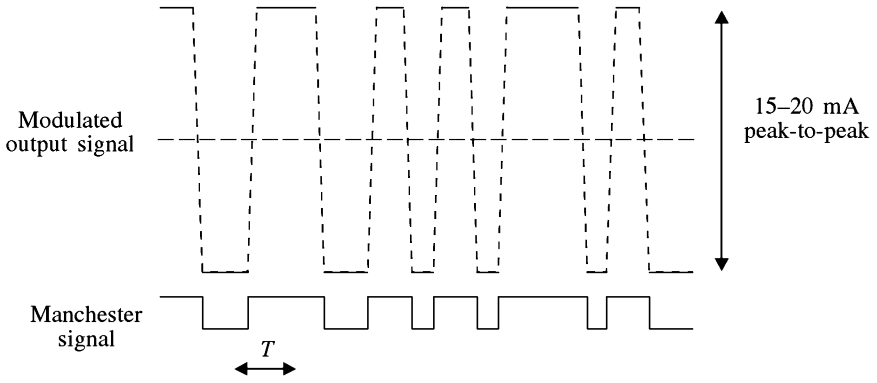


Figure 14.46 Wire media signal modulation.

The lowest speed wire media option, 31.25 Kbit/s, is the most versatile and expected to be the most widely used type. It offers versions for intrinsic safety and device powering by bus. The number of devices is limited by this choice.

- Intrinsically safe/nonintrisically safe
- Bus powered/separately powered

The typical number of devices is indicated in Table 14.13, the actual number varies from device type to device type.

Table 14.13 Maximum number of network nodes

	Intrinsically safe	*Non-intrinsically safe*
Bus powered	2-6*	2-12**
Separately powered		2-32

* 1–4 in hazardous areas

** 1 at power supply end

In intrinsically safe systems, the safety barrier is placed between the power supply and the power supply end terminator. This IS limit is partly because of the combination of several energy storage devices on the one barrier-limited bus and also to the combined power draw of several devices through one current-limiting resistor.

Devices are powered by the bus, only requiring two wires for supply and communication. A single power supply, common to all devices, is connected to the network at either end of the trunk. The voltage is in the range 9–32 VDC. The impedance of the power supply is a minimum of 3 kΩ around the transmission frequency in order not to short circuit the communication signal. A Manchester coded signal (explained below) has a duty-cycle of exactly 50% and can be seen as an ac signal. The dc power consumption (power drawn) of a device is therefore constant.

Communication

The data are interchanged as a synchronous serial *half-duplex* signal. A device transmits and receives on the same media, but not simultaneously. The signal is self-clocking, using Manchester (Biphase-L) encoding. Since the transmission is synchronous, no start and stop-bits are required. In Manchester coding every bit of data is a complete cycle. Clock and data are combined so that a rising edge (mid-cycle) represents logic 0 (zero) data, and a falling edge represents logic 1 (one) data as shown in Figure 14.47. The advantages of this relate to data security, in having every bit denoted by a change of state, and in system time-keeping.

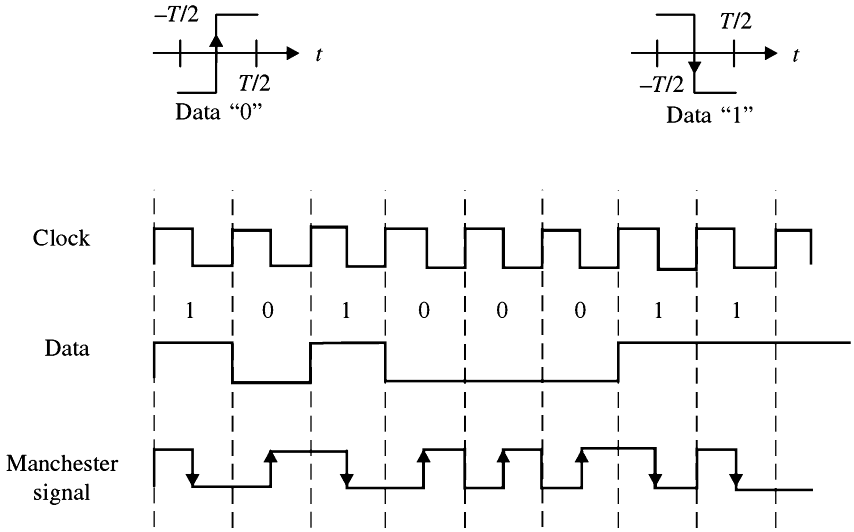


Figure 14.47 Manchester encoding.

Physical layer frame is shown in Figure 14.48. When transmitting, a *preamble* equivalent to a phone's 'ring' signal, is first transmitted to synchronize the receivers of other devices. The beginning and end of the message are indicated by start and end delimiters respectively. The delimiters are not Manchester encoded; only the data is, and can therefore be uniquely identified. The non-encoded bits in the delimiters are called N+ (non-data positive), and N–(non-data negative). The preamble and the delimiters, added by the physical layer in the transmitting device, are stripped off by the physical layer of the receiving device.

Redundancy may be achieved by duplication of physical layer and media. Voting (based on timeout and validity) controls which of the two media a device is using.

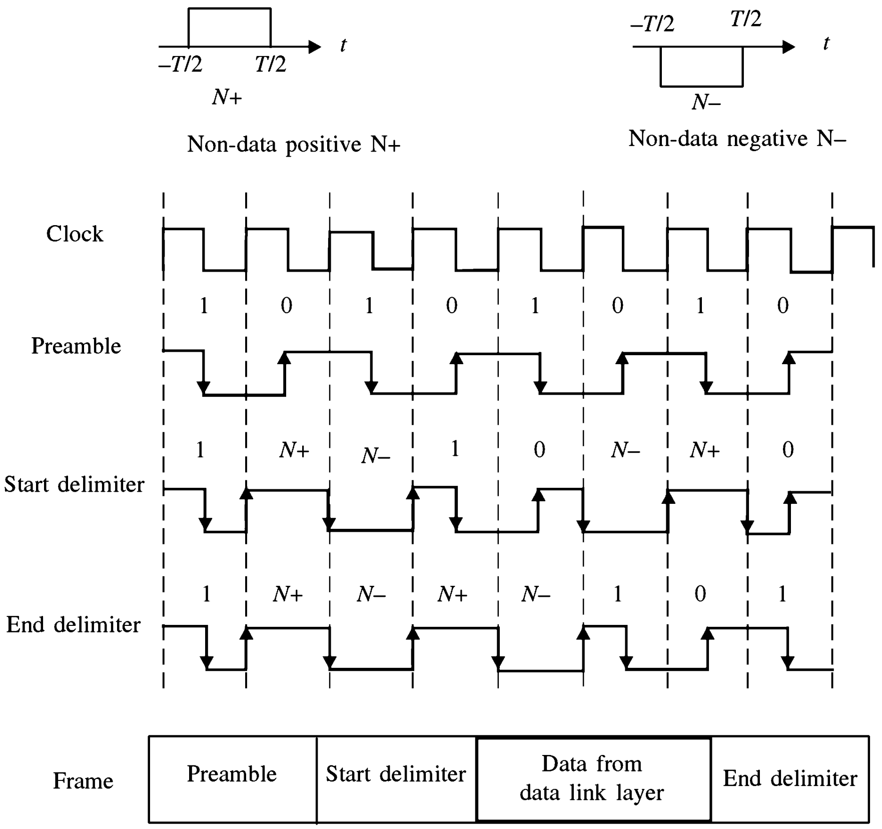


Figure 14.48 Fieldbus physical layer frame.

14.6.3 Data Link Layer

The Fieldbus data link layer follows a subset of the emerging IEC Fieldbus: Part 3 standard. Its function is to control transmission of messages to and from the fieldbus through the physical layer.

This layer consists of two sublayers: the lower portion is the Fieldbus Media Access Control (FMAC) and the upper portion is the Fieldbus Data Link Control (FDLC).

A device on the Fieldbus network is one of two types—Master or slave.

A master station has the right to access the media (initiate communication). Slaves only have the right to respond to a request from a master.

Fieldbus Media Access Control (FMAC)

The Fieldbus medium access is a fusion of the Token- passing and Polling principles. Several devices on a network may be master stations. Only the station that has the token is permitted to initiate communication. The master may poll (master requests, slave responds) the slave devices while it has the token. The token is passed to the next master in a special frame.

The devices are given individual station addresses. All frames contain the Destination Address (DA) and the Source Address (SA) for the message.

The Fieldbus has services that free the user from the responsibility of assigning and keeping track of addresses.

A requirement for reliable control is reliable data. A two-byte Frame Check Sequence (FCS) is calculated on all frame data using a polynomial in the transmitting device, and added to the frame. The receiving device performs the same calculation and compares the result with the FCS, thereby detecting any error. The FCS is equivalent to the parity-bits and Cyclic Redundancy Checks of asynchronous protocols.

When the above layer requests the FDL to pass a message, the message priority is passed along with it. There are two priorities: high, e.g. alarms, and low, e.g. configuration and diagnostics data. The FDL transmits the high priority messages first.

Fieldbus Data Link Control (FDLC)

The FDLC provides various possibilities for the application layer to send data to other stations.

There are two message types that can be identified in a Fieldbus system—Operational and Background.

Operational traffic is data transferred between devices as part of the control strategy, e.g. process variables. It is characterized as low volume, time critical and cyclic. Background traffic is data transferred between a device and the operator interface, e.g. configuration and diagnostics. It has the opposite characteristics of operational traffic—high volume, not time critical and is acyclic (sporadic).

Object oriented design

When dealing with a complex system such as a Fieldbus application, the whole may appear unfathomable. By decomposing it to parts, and even more primitive elements in a hierarchic nature, to a suitable level of abstraction (recognizing essential characteristics), complex interacting parts are brought to order and the system becomes easier to grasp. That is achieved since one only needs to comprehend a few simple parts. A technique, called object oriented design (OOD), was used to design the application layer and the function block application process.

There are many *keywords* in OOD. However, for Fieldbus study, Object and Class are sufficient. Objects are entities with a well defined behaviour. Systems are broken down to objects which can be said to be *parts of* the system. In OOD, software is based on objects that do things or change when one sends them messages or operates upon them. Therefore, OOD is not based on algorithms (execution steps). Objects often represent entities in the real world, e.g. a file. Objects may be classified according to their function and other properties they have in common. A class defines various *kinds of* objects. Unique properties of an object defines an *instance* of the class.

To order or rank abstractions, both objects and classes are organized in a hierarchy (levels of abstraction or complexity). One being built on top of the other, each level is understandable on its own. Inheritance is class hierarchy—a subclass (lower class) shares structure and behaviour of a superclass (higher class). Multiple inheritance is possible, and common. Aggregation is object hierarchies—objects are built from sub-objects.

For example, in a large control system one can find management system, supervisory system and field equipment—all parts of the control system. A field device may have sub-parts such as sensor, electronics and casing. In a system there may be a LD 302, which is a kind of pressure transmitter which is a kind of transmitter, which is a kind of field equipment. The LD302 inherits the properties and behaviour of the pressure transmitter class. Therefore, an operator familiar with pressure transmitters can operate the LD 302 in a matter of minutes, only having to learn the unique properties of the LD 302.

OOD yields smaller systems through this reuse of common mechanisms, aggregation and inheritance.

Models are used extensively in Fieldbus and in all engineering because they make abstraction, decomposition and hierarchical ordering easier.

OOD in Fieldbus

The Fieldbus control system has been decomposed down to the *simple variables* which is a suitable level of abstraction. Examples of those are float, integer and string. Simple variables are used on their own but also as parts of *data structures* such as the function block I/O parameters and function block links. Again these data structures are parts of the function block data type which is also a data structure.

Function blocks are part of the function block application process, which is part of the field device which finally is part of the system.

Variables may be classified in many ways:

- float, integer or string
- static or dynamic
- read or write
- and so forth

Variables is only one of many types of objects, but the most important, defined in Fieldbus. Since they exist in the AP, they are called *Application Process Objects* (APO).

14.6.4 Application Layer

The distributed application processes in the system need to communicate. The Fieldbus provides logical communication paths (channels) between the application processes. Various types of connections with various combinations of characteristics are available to meet the various communication needs. Several connections may exist simultaneously, enabling multivariable access.

The Fieldbus connections are modeled in two ways— client-server model and publisher-subscriber model.

The client-server model is used to describe acyclic data transfer. From a communication point of view a *client* is an AP which is using a remote AP's functionality. The remote AP is called the *server*. For example, if the operator console wants to read a tuning parameter in a controller in the field, the AP in the console is the client, and the AP in the controller is the server.

The publisher-subscriber model is used to describe cyclic data transfer. It is derived from the more common producer-consumer model. From a communication point of view a *subscriber* is an AP which is using a remote AP's functionality. The remote AP is called the *publisher*. The publisher is actually producing (publishing) data, a subscriber is consuming that (subscribing to) that data. For example, a transmitter is publishing a process variable which is consumed by a controller. The controller is publishing an output which is consumed by an actuator. Transmission is controlled by a third party, the *requestor*, which issues a request to the publisher to publish its data.

As mentioned in the OOD introduction, a Fieldbus system is broken down into variables. There is a set of services that lets an AP use the functionality of an AP in another device, such as getting the value of a variable or otherwise manipulate the object.

The primary intention of Fieldbus is to build the application using function blocks. This would be done in the Function Block Application Process (FBAP). However, within a Fieldbus device it is possible to have other types of APs, e.g. ladder logic or structured text, though no such definition has been made yet.

From a Fieldbus point of view, a device is not its hardware parts as they are to humans. For example, a pressure transmitter is not an assembly of pressure sensor, electronics and a housing, but a network node containing parameters. This network view is called the Virtual Field Device (VFD). A device (station) contains only one FBAP. The FBAP may contain several VFDs to device a device's application into individual loops to make it easier for the operator to overview.

The VFD is the Interface between protocol stack and function block AP. The VFD is the part of the real application that is visible and accessible through the network, the communication objects such as variables and blocks, etc.

Before a device can access communication objects (e.g. variables) in another device, it must first know which objects are available and their structure. Knowing the structure is important because there is no point in asking for a variable if you do not know how to interpret the answer, if it is, e.g. a float or integer. This information may be pre-configured or obtained from the communication partner. There are two types of such services—operational services to manipulate objects, and services for manipulation of their descriptive attributes

All objects (variables, etc.) have an index for easy reference. Every parameter in the system is uniquely identified by its index plus the connection. This is the method used to request data once a Fieldbus system is up and running. The user does not have to worry about keeping track of indexes and addresses. That is done by the network and may be totally transparent to the user, depending upon the type of user interface.

A man machine interface like a hand-held terminal needs information about the object more than just for communication purposes. For example, it must know how to present the information to the user (e.g. in a menu), when it must be updated dynamically, if there is a certain procedure involved before writing the variable. This information may be stored in the device, or may be supplied separately on magnetic media, for example.

Examples of services are:

- release connection
- read device status
- read device manufacturer, type and version
- read entire or part of configuration
- read variable
- write variable
- notify an event
- instantiate block ('create')
- find OD index for a parameter
- delete block

System and network management

The purpose of network management is to provide services for central configuration and control of the network protocol stack, such as maintenance and start-up of the Fieldbus system. For example, the network management manages the connections. The system management is split into two parts—a kernel which provides the basic functionality that a control application can be built upon, and a part that provides optimization of operation and diagnostics of problems. It coordinates functions in all layers, controls overall device operation and start-up.

The system management kernel provides functions for:

- Device tag assignment
- Station address assignment
- Clock synchronization
- Scheduling of distributed APs
- Function block binding

In a system, each of the above functions can only be managed by one device (though one device may handle many of them), the others act as agents. In case a manager fails, one of the agents will assume the manager role. For the system management to perform its task, it must cooperate with the system management in other stations on the network. A simple device may implement only a part of the system management functions.

Physical device tag assignment. Before a device is put on the network, the user must first assign a physical device tag to the device (i.e. it is done off-line). The tag may be up to 16 characters, typically in accordance with the normal instrumentation practices, e.g. PT-10270.

Station address assignment. Automatically assigns and ensures that each device on the network has a unique address. An *uninitiated* device has a default address. Configuration devices detect new devices and will assign a station address, after check for duplicate tags, bringing the device to the *standby* state. A temporary device such as a hand-held configurator selects its own address if there is no traffic on the network.

Function block binding. The network automatically finds the device (station address) for a given function block. It checks for multiple tags. This function is used when resolving links between block outputs to inputs (identified by the block's tag, and the parameter name) to the short address and index reference.

Clock synchronization. For the Fieldbus system to perform scheduling and other time-related functions such as time stamping of alarms and events, there is a distributed time base (clock) in each device, providing a common sense of time among all devices—*system time*. System management provides a mechanism for synchronization of the time in each device. This is done from a *master clock* which provides the correct time.

Scheduling. The purpose of scheduling is to minimize delays due to communication. Such delays are pure dead time which make control difficult. Scheduling also ensures that variables are sampled and function blocks are executed on a precisely period basis, so that the delay is constant. A constant delay is a must, since a change in delay would require re-tuning. This way, tight closed-loop control is achieved with time left over for background traffic. Scheduling also ensures accurate trending and predictable alarm detection. There are three scheduled functions:

- Background traffic
- Operational traffic
- Function block execution

The synchronization is based on system time. A *macro cycle* is a period of block execution which is divided into an integer number of phases—*clock units*. All function blocks are executed, and all operational traffic is passed during the macro cycle (see Figure 14.49). The Function Block Execution starts the execution of the function blocks at the beginning of a phase by informing the FBAP at the proper time. Note that transducer blocks are not

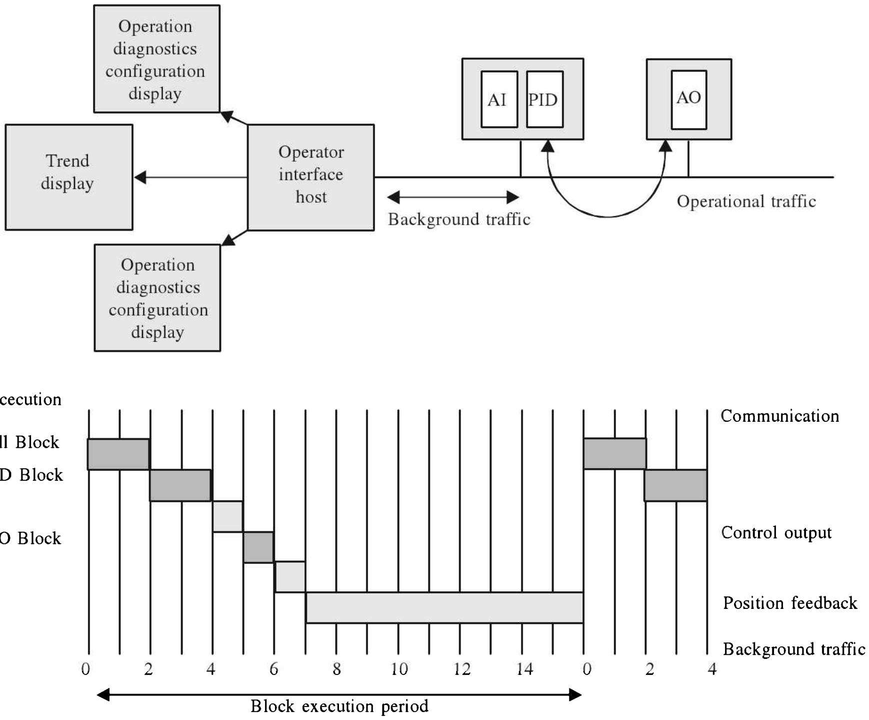


Figure 14.49 Scheduling.

scheduled to enable special sensing techniques or measurement when a sample is available. The user may determine in what order the blocks should be executed in order to minimize delays due to parameter propagation. Background Traffic is scheduled to be passed during the phases not used for Function block execution or operational traffic.

Execution period/time is given as follows:

$$\text{Execution time} = 2^x \text{ seconds}$$

where x is a signed integer between –128 and 127. For example. $x = -3$ results in 125 ms.

In the example in Figure 14.49, a simple control loop consists of input (Al), control (PID) and output (AO), with the Al and PID in the same physical device. No interdevice connection is necessary for the process variable. In this example, the execution period for the AO block is shorter in order to illustrate that block execution time depends on block.

Communication is scheduled in the master device which controls traffic and requests communication. Function blocks are scheduled in the individual device.

Function block application process

The Function Block AP (FBAP) is where the users configures their measurement and control application. Part of it is distributed to the various devices in the field. It is not executed in a single control card as it is done in a DCS.

The functionality of a Fieldbus device is modeled as objects (see Figure 14.50). The block object has three classes which again have sub-classes under which the various blocks are grouped.

- Block object
  - — Function block object
    - ♦ Input function block
    - ♦ Output function block
    - ♦ Control function block
    - ♦ Calculate function block
  - — Transducer block object
    - ♦ Input transducer block
    - ♦ Output transducer block
    - ♦ Display transducer block
  - — Physical block object
- Alarm object
- Event object
- Trend object
- Display list

The part of the FPAP, which is standardized by Fieldbus, is called the *function block shell.* For example, the block algorithms are not standardized. For each block there is a set of parameters that to a certain extent defines what minimum functionality a block will have. However, the manufacturers may implement such a block in their own way. For example, in the PID control block there must be a GAIN parameter and therefore, the manufacturer may use this parameter as gain or proportional band.

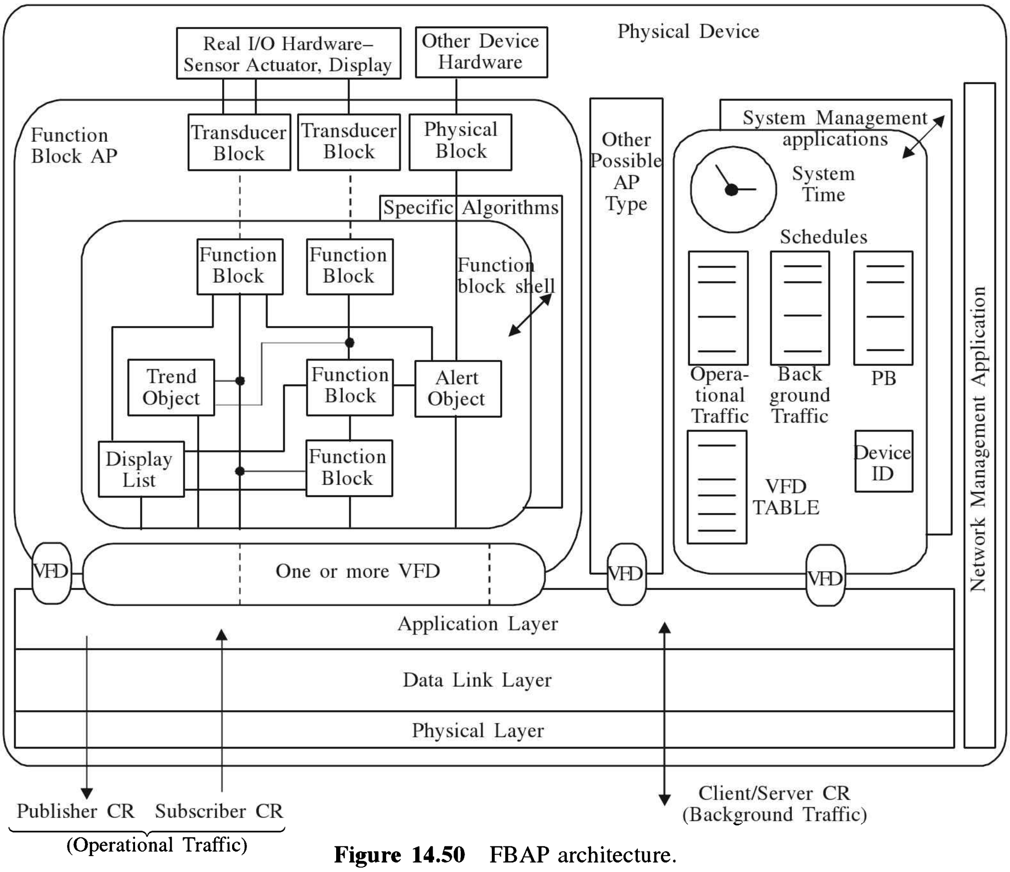


Figure 14.50 FBAP architecture.

Example

Let us use various function blocks to make the most simple measurement using only a single analog input block and to build classic control strategies like single loop, cascade, ratio, cross limit, etc. These are shown in Figure 14.51 through Figure 14.53.

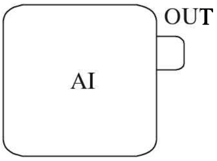


Figure 14.51 Analog measurement.

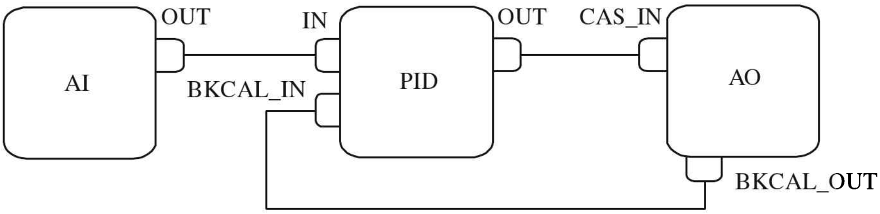


Figure 14.52 Single loop.

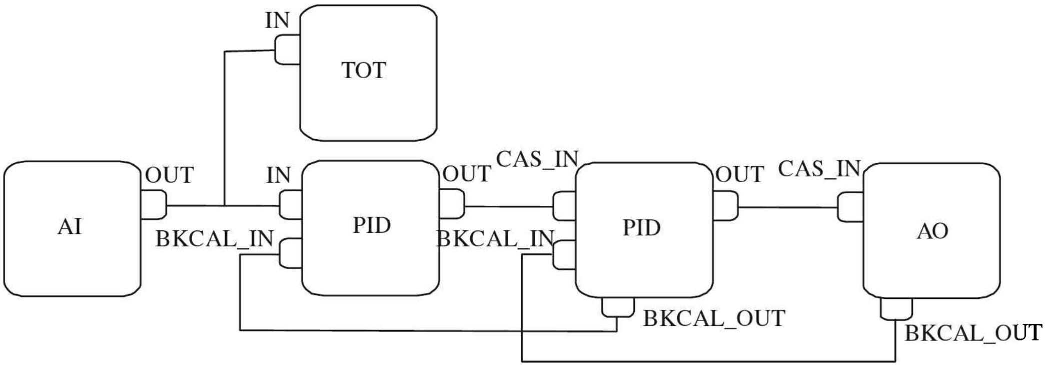


Figure 14.53 Cascade control.

The most basic input and output parameters are explained below:

IN: Process variable input.

OUT: Primary output of the block.

CAS_IN: Input for remote setpoint from another block cascaded setpoint.

BKCAL_IN: Feedback input from downstream block. Value used for initialization, balancing, of block output to ensure bumpless setpoint transfer in downstream block when it returns to Local mode.

BKCAL_OUT: Copy of selected setpoint to be used by upstream block for initialization of its output.

RCAS_IN: Input for setpoint from a host, selected in Remote cascade mode.

RCAS_OUT: Copy of the selected setpoint, value identical to BKCAL_OUT, but status is based on RCAS_IN communication.

ROUT_IN: Input for remote output from host, selected in Remote output mode.

ROUT_OUT: Copy of the selected output, OUT, but with status based on ROUT_IN communication.

SP (contained): Setpoint set by operator in local mode, or setpoint after limitation of setpoint received from other block or a host.

14.6.5 FF Benefits

Until real plants have been operating on Fieldbus equipment for several years, no one will truly know the full spectrum of benefits to accrue from the use of fieldbus. Further, whatever benefits are initially recognized and measured will quickly be outdated as the technology progresses. Here is a brief overview of some immediate benefits:

Lower installed equipment costs

- multidrop connection
- shared I/O interfaces
- fewer safety barriers
- added functions in field devices, e.g. PID

Lower maintenance costs

- self-reporting diagnostics
- capability/potential for predictive maintenance
- easier ISO 9000 calibration tracking
- less routine calibration with digital devices

Operational savings

- higher plant performance/stability with digital devices
- less downtime or reject-quality material because of faster response to problems
- more control data to take best fine-tuning decision or corrective action in case of upset

QUESTIONS

14.1 Draw the hierarchy of a factory automation. On what factors is an industrial communication system selected?

14.2 Explain the OSI network model.

14.3 Differentiate between ring, bus and star topologies.

14.4 Explain the RS 232 standard.

14.5 (a) What are the limitations of RS 232?
(b) What is bluetooth technology? Explain its features.

14.6 Explain RS 422, RS 423, and RS 425 standards.

14.7 Differentiate between RS 232, RS 422, RS 423, and RS 425 standards.

14.8 Explain IEEE 488 (GPIB) interface.

14.9 In what configurations are various devices connected in GPIB interface? What type of cables can be used for interconnection?

14.10 Compare IEEE 488.1 with IEEE 488.2 interface.

14.11 Explain HS 488 protocol.

14.12 Explain HART field communication protocol. What are the benefits of HART?

14.13 Explain the functions of a bridge, repeater and router.

14.14 With examples, explain the need of a Fieldbus.

14.15 Compare Direct Digital Control, Distributed Control and Fieldbus systems.

14.16 Name a few Fieldbuses, their developers and their basic applications.

14.17 Explain the architecture of the Foundation Fieldbus.

14.18 What are the topologies supported by the Foundation Fieldbus? Draw their connections.

14.19 What is the Function Block Application Process in the Foundation Fieldbus? Explain it with some examples.

14.20 What are the benefits of Foundation Fieldbus?

Chapter 15

Instrumentation in Hazardous Areas

15.1 INTRODUCTION

Hazardous areas are those locations where a potential for explosion and fire exists because of flammable gases, vapours or finely pulverized dusts in the atmosphere, or because of the presence of easily ignitable fibers or flyings. Such environments are found in chemical plants, petroleum refineries and other industries where combustible liquids, gases, or vapours exist.

Hazardous areas may result from the normal processing of certain volatile chemicals, gases, grains, etc. or they may result from accidental failure of storage systems for these materials. It is also possible that a hazardous location will be created when volatile solvents or fluids, used in a normal maintenance routine, vaporize to form an explosive atmosphere.

In hazardous areas, fire or explosion could occur when all the following three basic conditions are fulfilled:

(a) Combustible materials like flammable gas, vapour, dust or fibers are present.
(b) The combustible materials are mixed with air in the proportions required to produce a flammable mixture.
(c) A source of ignition acts to ignite the mixture.

Figure 15.1 shows the ignition triangle.

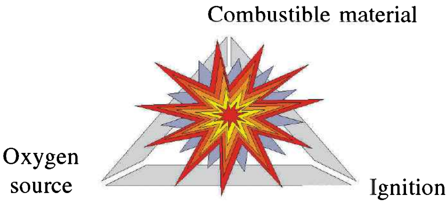


Figure 15.1 Ignition triangle.

The ignition source may occur in any of the four mechanisms: (a) discharge of capacitive circuits, (b) interrupting (opening) of inductive circuits, (c) opening or closing of resistive circuits with slow intermittent interruption increasing the ignition capability (hazard), and (d) high temperature sources. Therefore, the normal operation of switches, circuit breakers, motor starters, contactors and plugs, and receptacles can release this energy in the form of arcs and sparks as the contact open and close, i.e. making and breaking circuits. Lighting fixtures and motors are classified as *heat producing*, and they will become a source of ignition if they reach a surface temperature which exceeds the ignition temperature of the particular gas, vapour or dust in the atmosphere.

It is also possible that an abnormality or failure in an electrical system could provide a source of ignition. A loose termination in a splice box or a loose lamp in a socket can be the source of both arcing and heat. The failure of insulation from cuts, nicks or aging can also act as an ignition source again from sparking, arcing and heat. Other possible sources of ignition are electrostatic effects and frictional sparking.

In explosive atmospheres, the instrumentation system must be configured to prevent or control system-caused ignition of the atmosphere. A number of standard methods of protection against ignition have been established and these have been codified in construction standards. These design codes enable manufacturers to make equipment of a uniform type and have it tested by certification authorities for compliance with the standards.

Three basic approaches are used individually or in combination to prevent ignition of hazardous atmospheres:

(a) *Intrinsically safe/non-incendive equipment.* Eliminates the ignition source by designing equipment that cannot cause ignition when in contact with flammable or explosive atmospheres.
(b) *Gas purge/hermetic sealing.* Controls the atmosphere surrounding the potential ignition source so that components are enveloped within a non-hazardous atmosphere.
(c) *Flameproof/explosion-proof housing.* Contains the explosion within a housing, preventing the ignition of the surrounding atmosphere.

Hazardous areas have been classified into classes, divisions, zones, and groups by many organizations through different standards. Accordingly the equipment to be used in these areas has to follow these norms and must specify the category under which it can be used. In this chapter, we will study the classifications used and procedures employed for the equipment/instrumentation to work in the explosive locations.

15.2 HAZARDOUS AREA CLASSIFICATIONS

The classifications have been standardized by organizations of different countries. In the United States, the National Electrical Code (NEC), Article 500 classifies the areas by Divisions, Classes and Groups, and Article 505 classifies the areas by Zones and Groups. In Canada, Canadian Electrical Code(CEC), Section 18 of CSA code classifies the areas by Zones and Groups. In Europe, CENELEC (European Committee for Electrotechnical Standardization) and IEC (International Electrotechnical Commission) classify the areas by Zones and Groups.

Table 15.1 tabulates the classifications of hazardous locations.

Table 15.1 Hazardous locations classifications

Classification	*IEC, CENELEC, NEC 505 codes*	*NEC 500 CSA codes*
Material Presence		
Continuously present or present for long periods (typically >1000 hours per year)	Zone 0 (gases) Zone 20 (dusts)	Division 1
Intermittently present or present in normal operation (typically >10 and < 1000 hours per year)	Zone 1 (gases) Zone 21 (dusts)	Division 1
Abnormally present only for short periods (typically < 10 hours per year)	Zone 2 (gases) Zone 22 (dusts)	Division 2
Apparatus		
Gas and Vapours		
Acetylene	Group IIC	Class I/Group A
Hydrogen	Group IIB	Class I/Group B
Ethylene	Group IIB	Class I/Group C
Propane	Group IIA	Class I/Group D
Methane	Group I	N/A
Dust		
Metal	N/A	Class II/Group E
Coal	N/A	Class II/Group F
Grain	N/A	Class II/Group G
Fibers (All)	N/A	Class III

Classification	*IEC, CENELEC, NEC 505 codes*	*NEC 500 CSA codes*
Max Surface Temp. (°C)		
450	T1	T1
300	T2	T2
280		T2A
260		T2B
230		T2C
215		T2D
200	T3	T3
180		T3A
165		T3B
160		T3C
135	T4	T4
120		T4A
100	T5	T5
85	T6	T6

Some typical locations are:

Class I

- Petroleum refineries, and gasoline storage and dispensing areas
- Dry cleaning plants where vapours from cleaning fluids can be present
- Spray finishing areas
- Aircraft hangars and fuel servicing areas
- Utility gas plants, and operations involving storage and handling of liquified petroleum gas or natural gas.

Class II

- Grain elevators
- Flour and feed mills
- Plants that manufacture, use or store magnesium or aluminum powders
- Producers of plastics, medicines and fireworks

- Producers of starch or candies
- Spice-grinding plants, sugar plants and cocoa plants
- Coal preparation plants and other carbon handling or processing areas.

Class III

- Textile mills, cotton gins
- Cotton seed mills, flax processing plants
- Plants that shape, pulverize or cut wood and create sawdust or flyings.

Group resistivities and breakdown voltages

In each group, resistivities of the dust and breakdown voltages vary. Group E dusts are those having resistivities lower than 100 ohm-cm or which breakdown when subjected to 1000 volts/cm across a bulk sample. This group includes dust of metals, generally regarded as conductors. Group F dusts are those having resistivities between 100 ohm-cm. and 100 Mohm-cm and which do not breakdown when subjected to 1000 volts/cm across a bulk sample. This group includes the carbonaceous dust, generally regarded as semiconductors. Group G dusts are those having resistivities greater than 100 Mohm-cm and which do not breakdown when subjected to 1000 volts/cm across a bulk sample. This group includes the agricultural and plastic dusts generally regarded as insulators.

15.2.1 Protective Concepts

Where potentially explosive atmospheres are present, equipment and systems need to employ a method of protection to prevent ignition of external gases or dusts. Table 15.2 gives commonly used methods of protection.

Table 15.2 Protective methods

Protection method	*European (CENELEC)*	*IEC code*	*Zone*	*N. American practice*	*Class /Division*	*General principle*
Flameproof	EEx d	Ex d	1, 2	Explosion-proof	I/(1, 2)	Contain the explosion, quench the flame
Powder-filling	EEx q	Ex q	1, 2	Not recognized	–	
Pressurisation	EEx p	Ex p	1, 2	Purging	I/(1, 2)	Keep flammable gas out
Encapsulation	EEx m	Ex m	1, 2	Not recognized	–	
Oil-immersion	EEx o	Ex o	1, 2	Oil-immersion	I/(1, 2)	
Increased safety	EEx e	Ex e	0, 1, 2	Not recognized	–	No arcs, sparks or hot surfaces
Intrinsic safety	EEx ia	Ex ia	0, 1, 2	Intrinsic safety	I/(1, 2)	Limit energy of sparks and surface temperature
	EEx ib	Ex ib	1, 2	Not recognized	–	

EExia or Exia means intrinsically safe with two fault conditions and EExib or Exib stands for intrinsically safe with one fault condition. Ex representation scheme is an international system of certification for explosion protected apparatus administered by the IEC and described by IECEE-04. The eventual goal of this scheme is that a manufacturer of hazardous location electrical equipment would be able to obtain a single ‘Ex’ Certificate of Conformity from one Certification Laboratory and sell that product in any participating country without any legal or technical obstacle and without the need to get it recertified locally.

Figure 15.2 gives the hazardous area and equipment protection reference scheme.

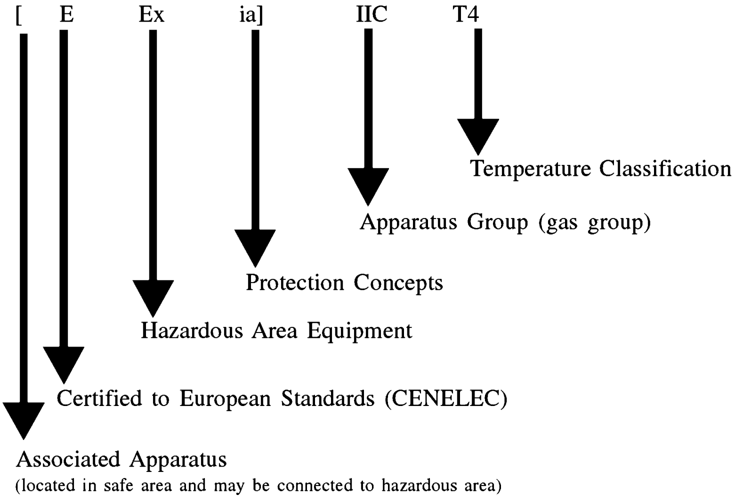

[E Ex ia] IIC T4
Temperature Classification
Apparatus Group (gas group)
Protection Concepts
Hazardous Area Equipment
Certified to European Standards (CENELEC)
Associated Apparatus
(located in safe area and may be connected to hazardous area)

Figure 15.2 Hazardous area and equipment protection reference scheme.

Table 15.3 summarizes ignition temperature and ignition energy levels of class II materials. Here, associated electrical apparatus may be either: (a) electrical apparatus that has an alternative type of protection for use in the appropriate hazardous (classified) location, or (b) electrical apparatus that is not protected and, therefore, cannot be used within a hazardous (classified) location.

Table 15.3 Ignition temperature and ignition energy of Class II materials

Group	*Representative material*	*Autoignition temperature* (°C)	*Ignition energy* (mW/s)
A	Acetylene	305	0.017
B	Hydrogen	520	0.017
C	Ethylene	450	0.08
D	Methane	630	0.3
E	Aluminum dust	550	15.0
F	Hard coal Kentucky bituminous	180	60.0
G	Wheat	220	240.0

Any product to be used for hazardous location must be approved by the concerned authorities of that country. It must be designed such that an explosion of the flammable or combustible material surrounding the device does not occur. Two most accepted methods are *explosion proof* (flame proof) enclosures and *intrinsic safety*.

Explosion-proof/flame-proof enclosures

An explosion-proof (or flame-proof, as classified in IEC and CENELEC standards) device is an electrical device designed with an enclosure capable of withstanding, without damage, an explosion within it of a specific gas, fiber, or dust. In turn, it prevents ignition of these same materials surrounding the enclosure by a spark or flame from the explosion within. This certification usually requires that devices be designed with sturdy and durable enclosures with conduit connections. The primary benefits of this type of protection are that the device is not limited by low available power nor does it restrict PCB layout as in the case of intrinsically safe devices.

Intrinsic safety

An intrinsically safe piece of equipment is an electrical device that is incapable of causing an ignition of the prescribed flammable gas, vapour, or dust, regardless of any spark or thermal effect that may occur in normal use, or under any conditions of fault likely to occur in practice. This means that the device design is limited in such areas as PCB layout, surface temperature, protection of electrical components, and power supply to the device. The devices are certified with either specific intrinsic safety barriers (loop certification) or general intrinsic safety barrier parameters (entity certification). These barriers are used outside the hazardous location and limit the amount of current, voltage, capacitance, and inductance entering the certified device. Often considered the safest and most technically elegant approach, there are many benefits of an intrinsically safe device. Expensive and cumbersome explosion-proof enclosures and conduit connections are not needed, electric shock is minimized, and controls can be maintained without shutting down the process.

We will first discuss the classification of enclosure designations and then the details of the intrinsic safety technique.

15.3 ENCLOSURE CLASSIFICATION DESIGNATIONS

An enclosure is a surrounding case constructed to provide a degree of protection to personnel against incidental contact with the enclosed equipment and to provide a degree of protection to the enclosed equipment against specific environmental conditions.

Both IEC publication 60529, developed by Electrotechnical European Standardization Committee (CENELEC), and DIN Standard number 40050, address the classification of degrees of protection provided by enclosures. This classification is designated as Ingress Protection (IP) rating (or code). These standards, however, do not specify degrees of protection against mechanical damage of equipment, risk of explosions, or conditions such as moisture (produced for example by condensation), corrosive vapours, fungus, or vermin.

The NEMA (National Electrical Manufacturer's Association) standard (NEMA 250-1997) for enclosures does test for environmental conditions such as hazardous conditions, corrosion, rust, icing, oil, and coolants. This classification is designated as NEMA types.

The following is a brief overview of the coding system described in these standards.

15.3.1 Ingress Protection (IP) Rating (or Code)

This consists of the letters IP followed by two numerals. The first characteristic numeral denotes the level of protection against solid objects and the second characteristic numeral indicates the level of protection against liquids (basically harmful ingress of water). For example, in the code IP 54, IP identifies this standard, the 5 describes the level of protection from solid objects, and 4 describes the level of protection from liquids. Table 15.4 gives the meaning of the first and second numerals in this code.

Table 15.4 The meaning of the first and second characteristic numerals in the IP code

First characteristic numeral	*Degree of protection (first number in code)*	
	Brief description	*Definition*
0	Not protected	—
1	Protected against solid foreign objects of 50 mm diameter and greater.	The object probe, sphere of 50 mm diameter, shall not fully penetrate.[1]
2	Protected against solid foreign objects of 12.5 mm diameter and greater.	The object probe, sphere of 12.5 mm diameter, shall not fully penetrate.[1]
3	Protected against solid foreign objects of 2.5 mm diameter and greater.	The object probe, sphere of 2.5 mm diameter, shall not penetrate at all.[1]
4	Protected against solid foreign objects of 1 mm diameter and greater.	The object probe, sphere of 1 mm diameter, shall not penetrate at all.[1]
5	Dust-protected	Ingress of dust is not totally prevented, but dust shall not penetrate in a quantity to interfere with satisfactory operation of the apparatus or to impair safety.
6	Dust-tight	No ingress of dust.

[1] The full diameter of the object shall not pass through an opening of the enclosure.

(*Contd.*)

Table 15.4 The meaning of the first and second characteristic numerals in the IP code (*Contd.*)

Second characteristic numeral	*Degree of protection (second number in code)*	
	Brief description	*Definition*
0	Not protected	–
1	Protected against vertically falling water drops.	Vertically falling drops shall have no harmful effects.
2	Protected against vertically falling water drops when enclosure is titled up to 15°.	Vertically falling drops have no harmful effects when the enclosure is tilted at any angle up to 15° on either side of the vertical.
3	Protected against spraying water.	Water sprayed at an angle up to 60° on either side of the vertical shall have no harmful effects.
4	Protected against splashing water.	Water splashed against the enclosure from any direction shall have no harmful effects.
5	Protected against water jets.	Water projected in jets against the enclosure from any direction shall have no harmful effects.
6	Protected against powerful water jets.	Water projected in powerful jets against the enclosure from any direction shall have no harmful effects.
7	Protected against the effects of temporary immersion in water.	Ingress of water in quantities causing harmful effects shall not be possible when the enclosure is temporarily immersed 1 metre in water under standardized conditions of pressure and time.
8	Protected against the effects of continuous immersion in water.	Ingress of water in quantities causing harmful effects shall not be possible when the enclosure is continuously immersed in water under conditions which shall be agreed between manufacturer and the user, but are more severe than for number 7.

15.3.2 NEMA Types

Table 15.5 summarizes the protection provided by various types of NEMA enclosures under different locations. Table 15.6 gives conversion of NEMA Type numbers to IEC Classification Designations (IP) and vice versa. This conversion, however, cannot be used to convert IEC Classification Designations to NEMA Type numbers.

Table 15.5 Protection provided by various types of NEMA enclosures

NEMA type number	*NEMA definition*	*Application for NEMA enclosures*
1	Enclosures are intended for indoor use primarily to provide a degree of protection against contact with the enclosed equipment in areas where unusual service conditions do not exist.	Enclosures are suitable for general application indoors, where atmospheric conditions are normal. These enclosures provide limited protection against falling dust, but are not dust tight.
2	Enclosures are intended for indoor use primarily to provide a degree of protection against limited amounts of failing dust and water. NEMA Standard 10-01-1979	Drip-tight (indoor) enclosures are similar to Type 1 enclosures, with the addition of drip shields, and are suitable for application where condensation may be severe, such as that encountered in cooling rooms or laundries.
3	Enclosures are intended for outdoor use primarily to provide a degree of protection against windblown dust, rain, sleet and external ice formation. NEMA Standard 10-01-1979	Suitable for applications outdoors on ship docks, canal locks, construction work and for application in tunnels and subways. Use indoors where dripping water is a problem.
3R	Enclosures are intended for outdoor use primarily to provide limited protection against falling rain, sleet and external ice formation. (May be ventilated). NEMA Standard 10-01-1979	Outdoor use to provide a degree of protection against falling rain and sleet; undamaged by the formation of ice on the enclosure.
4	Enclosures are intended for indoor or outdoor use primarily to provide a degree of protection against windblown dust and rain, splashing water, and hose-directed water. NEMA Standard 10-01-1979	Water tight enclosures are suitable for dairies, breweries, etc., where they are subjected to large amounts of water from any angle. (They are not submersible).
4X	Enclosures are for indoor or outdoor use primarily to provide a degree of protection against corrosion, windblown dust and rain, splashing water, and hose-directed water. NEMA Standard 10-01-1979	Corrosion resistant enclosures satisfy the requirements of Type 4, and are suitable for food processing plants, dairies, refineries, and any other industries where corrosion is prominent.
5	Enclosures are intended for indoor use primarily to provide a degree of protection against settling airborne dust, falling dirt and dripping non-corrosive liquids. NEMA Standard 25-05-1988	
6	Enclosures are for indoor or outdoor use primarily to provide a degree of protection against the entry of water, during occasional, temporary submersion at a limited depth. NEMA Standard 10-01-1979	Submersible enclosures are suitable for application where the equipment may be subject to submersion, such as quarries, mines and manholes. The enclosure design will depend upon the specified conditions of pressure and time.

(*Contd.*)

Table 15.5 Protection provided by various types of NEMA enclosures (*Contd.*)

NEMA type number	*NEMA definition*	*Application for NEMA enclosures*
7	Enclosures are for indoor use in locations that are classified as Class I, Groups A, B, C or D, as defined in the National Electric Code. NEMA Standard 10-01-1979.	Hazardous areas indoor classified as shown.
8	Enclosures are for indoor or outdoor use in locations that are classified as Class I, Groups A, B, C or D, as defined in the National Electrical Code. NEMA Standard 10-01-1979.	Hazardous areas indoor or outdoor classified as shown.
9	Enclosures are for indoor use in locations classified as Class II, Groups E, F, or G, as defined in the National Electrical Code. NEMA Standard 19-05-1986	Hazardous areas indoor enclosures classified as Class II, Group E, F, or G.
10	Enclosures are constructed to meet the applicable requirements of the Mine Safety and Health Administration (Bureau of mines)	Provides protection against atmosphere containing methane with or without coal dust.
12	Enclosures are intended for indoor use primarily to provide a degree of protection against dust, falling dirt and dripping non-corrosive liquids. NEMA Standard 10-01-1979.	Indoor uses to protect the equipment against fibers flyings, lint, dust, and light splashing, seepage, dripping, and external condensation of non-corrosive liquids.
13	Enclosures are intended for indoor use primarily to provide a degree of protection against dust, spraying of water, oil, and non-corrosive coolant. NEMA Standard 10-01-1979	Indoor enclosures intended primarily to house pilot devices such as limit switches, push buttons, selector switches, pilot lights, etc. and to protect these devices against lint and dust, seepage, external condensation, and spraying of water, oil, and coolant.

Table 15.6 Conversion of enclosure type numbers

NEMA to IP

NEMA enclosure type number	*IEC enclosure classification designation*
1	IP10
2	IP21
3	IP54
3R	IP24
4 and 4X	IP55
5	IP50
6 and 6P	IP67
12	IP52
13	IP54

(*Contd.*)

Table 15.6 Conversion of enclosure type numbers (*Contd.*)

IP to NEMA

First	*Second digit of IP classification*							
digit	0	1	2	3	4	5	6	7
0	–	–	–	–	–	–	–	–
1	–	–	–	–	–	–	–	–
2	1	1+2	1+2	1+2	1+2+3R	1+2+3R	1+2+3R	1+2+3R
3	1	1+2	1+2	1+2	1+2+3R	1+2+3R	1+2+3R	1+2+3R
4	1	1+2	1+2	1+2	1+2+3R	1+2+3R	1+2+3R	1+2+3R
5	5	1+2	12	1+2	3+13	4+4X	1+2+3R	1+2+3R
6	5	1+2	12	1+2	3+13	4+4X	1+2+3R	6

Explosion-proof enclosures

Explosion-proof enclosures are designed to contain an explosion, not prevent one. They are generally constructed of cast iron or aluminum, and provide limited access.

Pressurized (purged) enclosures

Purging technologies are frequently used when installing a control system in a hazardous environment. The control system enclosure is usually rated NEMA 4 or 4X and is fitted with a means of providing a continuous flow of inert gas or instrument air which pressurizes the enclosure and prevents the intrusion of explosive substances. A pressure switch is installed to monitor for enclosure pressurization. Enclosures installed in areas assigned a Class I, Division 2 rating, require a Type Z purge system where the pressure switch generates an alert when enclosure pressurization is not present. Enclosures installed in Class I, Division 1 areas require the additional protection provided by Type X purge systems. The loss of enclosure pressurization requires an interruption of all electrical power to the enclosure. Before electrical power can be reenergized, the purge system must perform a *rapid exchange purge* in which four volumes of air are passed through the enclosure.

Since heat is so destructive to electronics, normal tendency is to *throw a lot of cooling* into an enclosure. However, too much cooling can create condensate problems. Enclosure manufacturers are able to offer a variety of cooling options that can be grouped into one of the following three categories.

- **Thermal/convection cooling** is accomplished by heat being dissipated through the skin of the enclosure, using a heat-sink arrangement, or by some form of fan/filter system.
- **Near-ambient cooling** is achieved using active air-to-air heat exchangers. This is a medium-cost solution and should be considered when the outside ambient temperature is below the maximum operating temperature of the electronics to be protected.
- **Below-ambient cooling** can be accomplished by traditional air-conditioner systems, solid state air-conditioners, water-to-air heat exchangers, and vortex-style air coolers.

The use of thermoelectric cooling, or *solid-state-cooling*, is gaining popularity. These coolers are suitable for washdown, and feature a smaller footprint than conventional air conditioning systems. With no compressor, evaporator, or condenser, they result in less maintenance. Solid state cooling uses dissimilar conductors of doped Bismuth Telluride to replace *refrigerant*. The compressor is replaced by dc power, supplied from an internal ac to dc power supply, which *pumps* electrons from one semiconductor to the other. The use of heat sinks and fans, both inside and out, replaces the *condenser* fins and enhances the cooling effect to achieve the required cooling.

15.3.3 Equipment Design and Construction

Equipment is designed and constructed to be suitable for hazardous locations for a particular Class applications. Let us start with Class I applications.

The first requirement for a Class I enclosure is strength. The enclosure must be strong enough to contain an explosion within. The walls must be thick enough to withstand the internal strain. It has to be explosion-proof in case gas or vapours get inside. Secondly, it must function at a temperature below the ignition temperature of the surrounding atmosphere.

The equipment must also provide a way for the burning gases to escape from the device as they expand during an internal explosion; but only after they have been cooled off and their flames *quenched*. This escape route for the exploding gases is provided through several types of flame paths.

One type is the ground surface flame path. Here the surfaces are ground, mated, and held to a tolerance of 15 ten-thousandths of an inch. This permits gases to escape, but only after they have been sufficiently cooled, so they would not ignite the volatile surrounding atmosphere. Another kind of flame path is the threaded flame path. After an explosion, the gas travels out the threaded joint but as it does, it cools off. Exploded gases may also escape around the shafts of operators used in the enclosure. But, here again, close tolerances are used to quench the burning gas. Examples of two flame paths are shown in Figures 15.3(a) and (b).

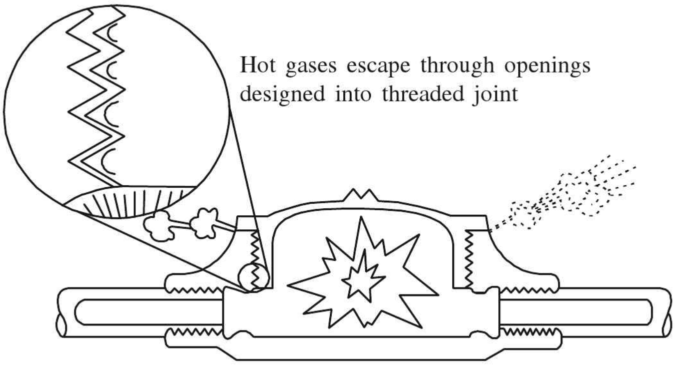

Hot gases escape through openings designed into threaded joint

Figure 15.3(a) An example of flame path: openings designed into threaded joint.

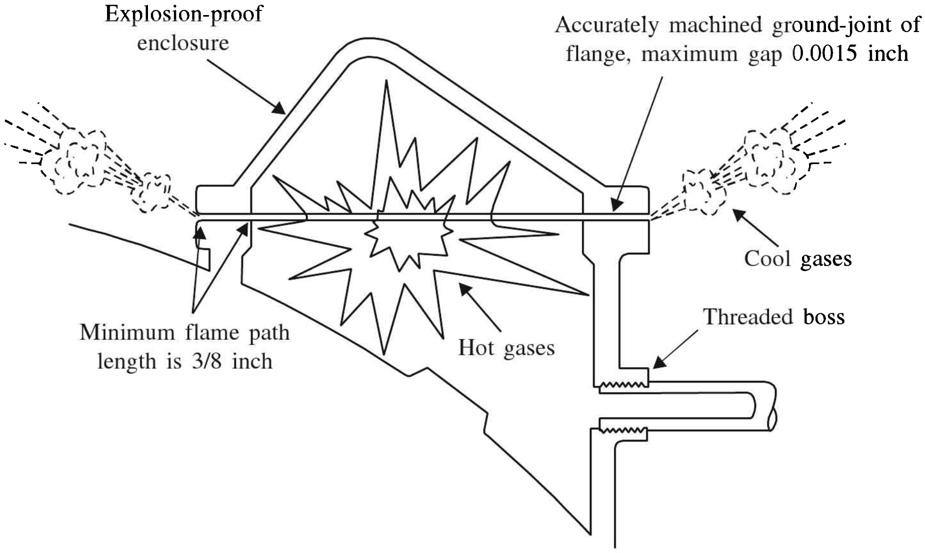

Explosion-proof enclosure
Accurately machined ground-joint of flange, maximum gap 0.0015 inch
Cool gases
Threaded boss
Minimum flame path length is 3/8 inch
Hot gases

Figure 15.3(b) An example of flame path: openings designed into ground joint.

We can see how important it is to make certain that all flame paths are protected during installation and maintenance, and even during handling, shipping, and storage of explosion-proof material. Even slight damage to a flame path can permit burning gases to escape, igniting the surrounding atmosphere. Also, all cover bolts must be installed for the same reason. A single missing bolt could allow the release of flaming gases.

In designing equipment for Class I, Division 1 locations, it is assumed that the hazardous gases or vapours will be present and eventually seep into the enclosure, so there is a real chance for an internal explosion to occur. In the case of Class II, however, the assumptions are different and so the design is different.

In Class II, the explosive dust is kept away from equipment housed within the enclosure so that no internal explosion can take place and there is no longer any need for heavy explosion-containing construction, or flame paths. This difference explains why Class I, Division 1 equipment can be called explosion-proof, and Class II equipment is called dust-ignition proof. Class II equipment has a different set of requirements:

(a) It must seal out the dust.
(b) It must operate below the ignition temperature of the hazardous substance.
(c) It must allow for a dust blanket, that is, the build-up of dust collecting on top of the device that can cause it to run *hot* and ignite the surrounding atmosphere.

For Class III equipment, there is very little difference in the design from Class II. Class III equipment must minimize entrance of fibers and flyings; prevent the escape of sparks, burning material or hot metal particles resulting from failure of equipment; and operate at a temperature that will prevent the ignition of fibers accumulated on the equipment.

There are many enclosures, devices, and fixtures suitable for all three classes. This simply means that it meets the specifications for each individual type. A Class I device which could contain an explosion of a specified gas would also have to prevent dust from entering the enclosure to be suitable for Class II. The close tolerance of the flame path which cools the burning gases is also close enough to exclude explosive dust so that a gasket would not be needed.

Proper installation of hazardous location equipment calls for the use of seals. Special fittings are required to keep hot gases from travelling through the conduit system igniting other areas if an internal explosion occurs in a Class I device. They are also needed in certain situations to keep flammable dusts from entering dust-ignition-proof enclosures through the conduit. As shown in the Figure 15.4, when arcs and sparks cause ignition of flammable gases and vapours, the equipment contains the explosion and vents only cool gases into the surrounding hazardous area.

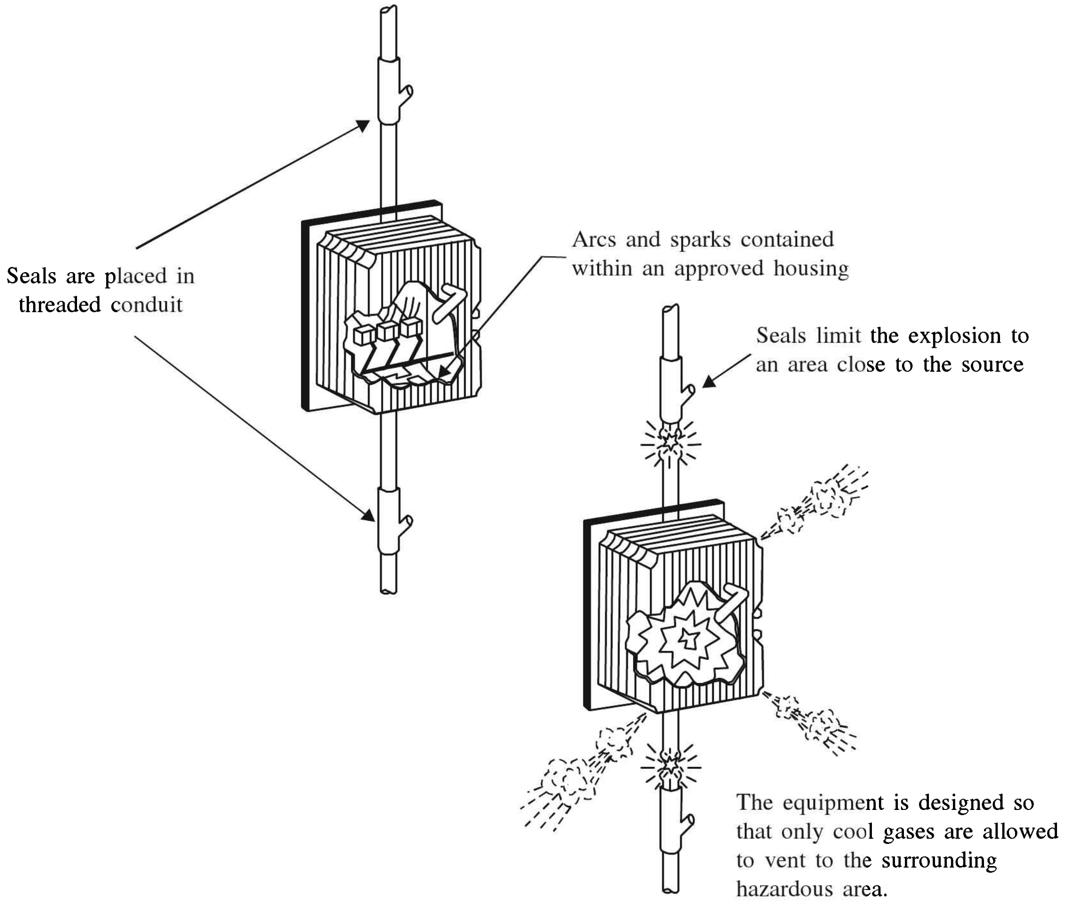

Seals are placed in threaded conduit
Arcs and sparks contained within an approved housing
Seals limit the explosion to an area close to the source
The equipment is designed so that only cool gases are allowed to vent to the surrounding hazardous area.

Figure 15.4 Use of seals.

Sealing fittings are designed to be filled with a chemical compound after the wires have been pulled. As the compound hardens, it seals passageways for dusts and gases. In each conduit run entering an enclosure for switches, circuit breakers, fuses, relays, resistors, or other apparatus which may produce arcs, sparks, or high temperatures within Class I locations, conduit seals shall be placed as close as practicable and in no case more than 18 inches (457 mm) from such enclosures.

Rigorous standards for hazardous location equipment have been set. Nationally recognized testing laboratories conduct actual explosion tests under laboratory conditions. For each Class I enclosure they experiment with different mixtures of gas and air, from very lean mixtures (a small percentage of gas) to very rich mixtures (a high percentage of gas) until they find the one that creates the greatest explosion pressure. To pass inspection, the equipment must not only prevent the ignition of the surrounding atmosphere, but also be able to withstand a hydrostatic test where oil is pumped into the enclosure at high pressure to test the limits of its strength. The device will not pass unless it can resist rupture at four times the maximum pressure found in the explosion tests. For example, if explosion testing shows a maximum pressure for a junction box of 250 pounds per square inch (psi), to get approval, the box must be able to withstand 1000 psi of hydrostatic pressure—four times the maximum anticipated pressure of 250 psi.

15.4 INTRINSICALLY SAFE DESIGN

Intrinsic Safety (IS) came into being around the turn of the century in the coal mining regions of England and has become a common protective technique universally accepted around the world. IS prevents instruments and other low-voltage circuits from releasing energy, in the form of sparks or heat, sufficient to ignite volatile gases in hazardous areas.

An intrinsically safe circuit not only allows instrumentation and control circuits to operate properly under normal conditions, it also protects them in the event of an electrical fault by preventing an ignition from sparks or overheating.

15.4.1 Determining Safe Energy Levels

Voltage and current limits are determined by ignition curves (Figure 15.5). A circuit having a combination of 30 V and 150 mA falls on the ignition level of some gases. Such a voltage/current combination could create a spark sufficient to ignite the right mixture of gases and oxygen. Intrinsically safe applications always remain below these threshold levels where the energy operating level is 1 watt or less. Capacitance and inductance curves must also be examined in intrinsically safe circuits.

Since intrinsic safety prevents instruments and other low-voltage circuits in hazardous areas from releasing sufficient energy to ignite volatile gases, the excess electrical energy in the form of voltage and current is limited by inserting energy-limiting devices, known as intrinsically safe barriers, in the circuits. The barriers designed to protect the system must be mounted outside the hazardous area, i.e. in an area designated as non-hazardous or safe in which the hazard is not and will not be present. The intrinsically safe barrier limits the current with a resistor and the voltage with a zener diode.

Products to be mounted in the hazardous area can be approved either under the LOOP or ENTITY approval concept. The LOOP concept specifies the exact part number and

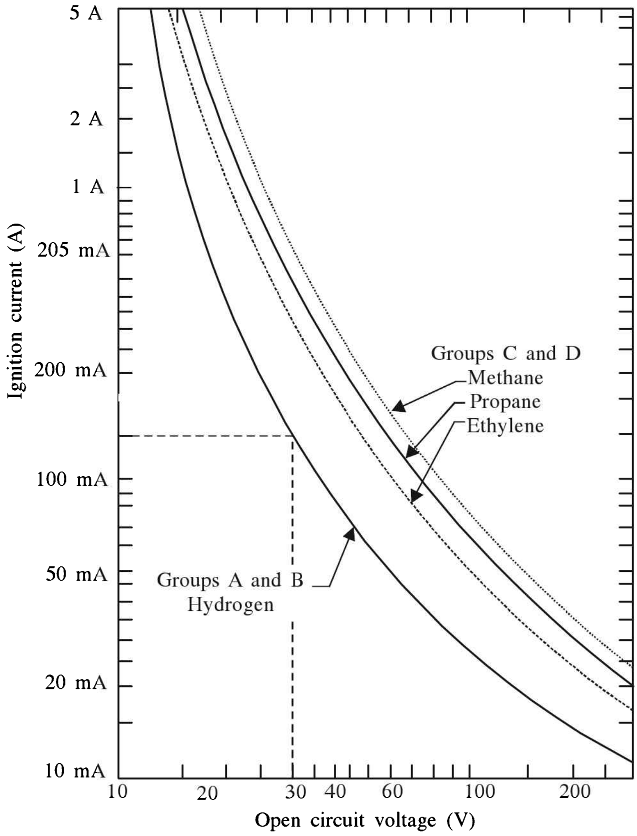


Figure 15.5 Ignition curves.

products that can be used in the loop. No deviation from the specified units is allowed. The ENTITY concept specifies parameters which any approved intrinsic safety barrier must meet. This allows the user to select barriers from different approved manufacturers. Under entity approval two items may be interconnected if the conditions specified in Table 15.7 are met.

Table 15.7 Conditions (entity values) for interconnection

Hazardous area (intrinsically safe approved apparatus)		*Non-hazardous (safe) area (intrinsically safe barrier)*
$V_{\max}$	must be less than	V_{oc}
$I_{\max}$	must be less than	I_{sc}
L_a	must be greater than	$L_i + L_w$
C_a	must be greater than	$C_i + C_w$

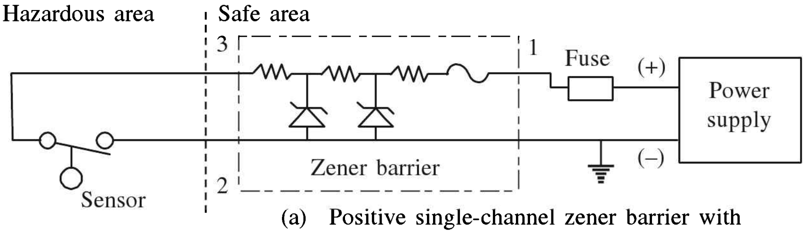
Here $V_{\max}$ = maximum open circuit voltage, $I_{\max}$ = maximum short circuit current, L_i = maximum unprotected inductance, C_i = Maximum unprotected capacitance, V_{oc} = maximum open circuit voltage (barrier), I_{sc} = maximum short circuit current (barrier), L_a = maximum allowed inductance (barrier), C_a = maximum allowed capacitance (barrier), L_w = inductance of interconnecting wiring, C_w = capacitance of interconnecting wiring. In all cases the intrinsically safe barriers and equipment must be wired as per an approved drawing. Capacitance and inductance of the wiring and cables must be included in the loop evaluation.

Table 15.8 summarizes IS parameter values of a typical Class IIC barrier.

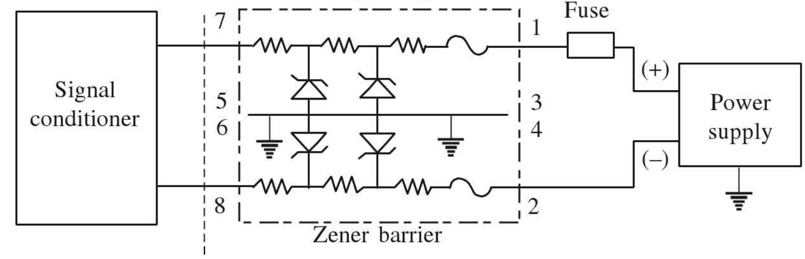
Table 15.8 IS parameter values of a typical Class IIC barrier

Abbreviation	*Parameter*	*Typical IIC barrier values*
$V_{\max}$	Maximum output voltage	22 V
$I_{\max}$	Maximum output current	214 mA
$P_{\max}$	Maximum output power	1.19 W
C_a	Maximum external capacitance	165 nF
L_a	Maximum external inductance	350 μH
L_a/R_a	Maximum external L/R ratio	31 μH/Ω

Figure 15.6 shows the use of two types of barriers, positive single-channel zener barrier with negative ground and positive dual-channel zener barrier with floating leads. Figure 15.7 explains how a barrier does not allow the energy flow to the hazardous area due to the short-circuit.



(a) Positive single-channel zener barrier with negative ground



(b) Positive dual-channel zener barrier with floating leads. *Note*: Terminals 3,4,5 and 6 are common and are bonded to the mounting tabs for positive redundant grounding.

Figure 15.6 Use of barriers.

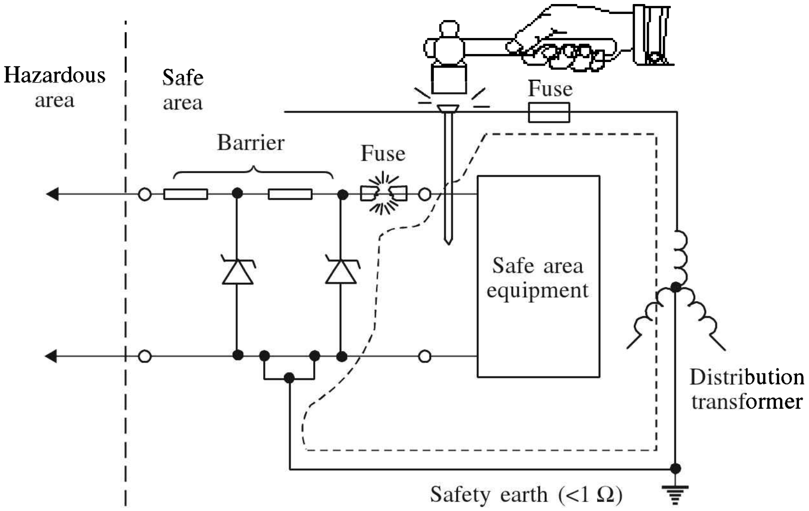


Figure 15.7 An example of a major failure and how barriers protect the area.

So we see that there are three components to an intrinsically safe circuit—the field device, intrinsically safe barrier, and field wiring. Field devices known as intrinsically safe apparatus are classified as simple or complex. Simple apparatus, which do not need to be approved, are non-energy storing devices such as contacts, thermocouples, RTDs, LEDs, and resistors. These simple devices do not need to be approved as intrinsically safe. If they are connected to an approved intrinsically safe associated apparatus (barrier), the circuit is considered intrinsically safe. Simple apparatus is defined in Paragraph 3.12 of the ANSI/ISA-RP 12.6-1987 as any device which will neither generate nor store more than 1.2 volts, 0.1 amps, 25 mW. Complex apparatus such as transmitters, solenoids, relays and transducers may store excess energy from the energy stated above and need to be approved by a third party. They must satisfy the conditions for interconnection as stated in Table 15.7. Contacts, transmitters and temperature sensors are the most commonly used field devices in intrinsically safe applications.

15.4.2 Intrinsic Safety Circuit Design

For designing an intrinsically safe circuit, we begin the analysis with the field device to determine the type of barrier that can be used so that the circuit functions properly under normal operating conditions but still is safe under fault conditions. More than 85% of all intrinsically safe circuits involve commonly known instruments. Table 15.9 shows the approximate use of intrinsically safe apparatus in hazardous areas.

There are three components to a barrier that limit current and voltage—a resistor, at least two zener diodes, and a fuse. The resistor limits the current to a specific value known as the short circuit current, I_{sc}. The zener diode limits the voltage to a value referred to as open circuit voltage, V_{oc}. The fuse will blow when the diode conducts. This interrupts the circuit, which prevents the diode from burning and allowing excess voltage to reach the hazardous area. There always are at least two zener diodes in parallel in each intrinsically safe barrier. If one diode fails the other will operate providing complete protection.

Table 15.9 Current use of intrinsically safe apparatus in hazardous areas

Intrinsically safe apparatus		*Intrinsically safe applications (%)*
Switching		32.0
mechanical switches	85.0%	
proximity switches	15.0%	
2-wire transmitters		22.0
Thermocouples and RTDs		13.0
Load cells		8.5
Solenoid valves		4.5
Potentiometers		2.5
LEDs		2.0
I/P transducers		2.0
Other devices		13.5

It also is important to make sure that the intrinsically safe circuit will work under normal conditions. With the current-limiting resistor, a voltage drop will occur between the input and output of the barrier. This has to be accounted for in the circuit design.

Is a thermocouple intrinsically safe?

A thermocouple is classified as a simple device. It does not create or store enough energy to ignite any mixture of volatile gases. If the energy level of a typical thermocouple circuit was plotted on the ignition curve (refer to Figure 15.5), it would not be close to the ignition levels of the most volatile gases in Group A. Then, is the thermocouple installed in a hazardous area (Figure 15.8) intrinsically safe? The answer is no, because a fault could occur on the recorder which could cause excess energy to reach the hazardous area (Figure 15.9). A barrier that limits the energy (Figure 15.10) must be inserted to make sure that the circuit remains intrinsically safe.

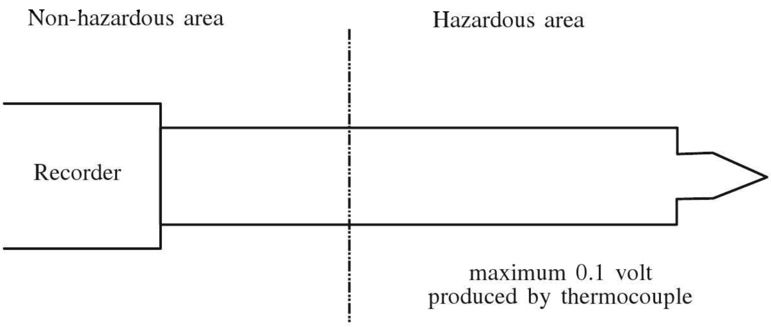


Figure 15.8 A thermocouple installed in an hazardous area.

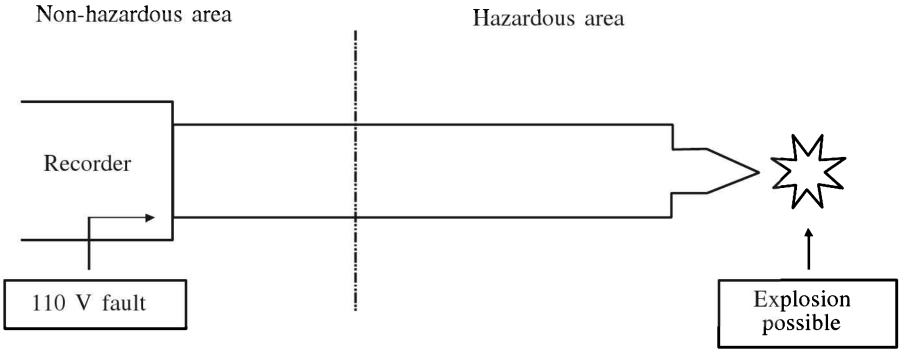


Figure 15.9 A 110 V fault reaching the thermocouple producing a spark in an hazardous area.

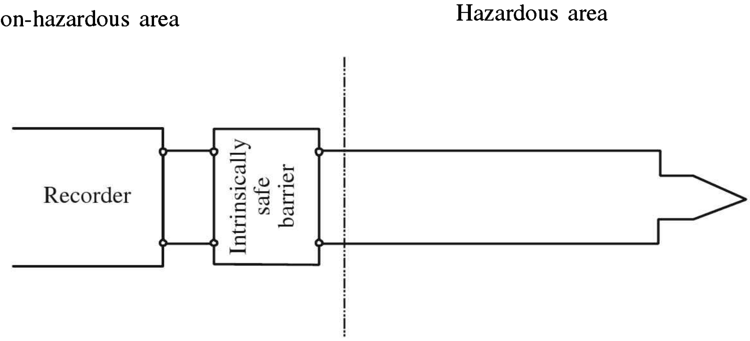


Figure 15.10 Using an intrinsically safe barrier.

Selection of a barrier

The parameters to select a barrier are—polarity of the circuit, rated voltage of the barrier, and resistance of the barrier.

Polarity. The circuit's polarity must be known in order to choose the correct type of barrier. The dc barriers are rated either as positive or negative. The ac barriers can be connected to circuits with either a positive or negative supply. SIGNAL and RETURN barriers are used for transmitter and switching applications. All of these barriers are available in single- or double-channel versions. However, because double-channel barriers save space and money by being connected to two legs of a loop, they are becoming the standard.

Rated voltage. Like any electrical device, safety barriers have a rated nominal voltage, V_n, referred to as working voltage. The barrier's V_n should be greater than or equal to the supply to the barrier, much like the rated voltage of a lamp must be equal to or greater than the supply to it. If the voltage supply to the barrier is much greater than its V_n, the barrier will sense a fault. The protective zener diodes will conduct, causing leakage currents and inaccurate signals on the loop. Most barriers have a rated working voltage that guarantees a minimal leakage current from 1–10 μA if the voltage is not exceeded. If the supply voltage to the barrier becomes too high, the zener diode will conduct. The resulting high current through the fuse will cause the fuse to blow. Excess supply voltage is the main reason why grounded barriers fail.

Internal resistance. Every safety barrier has an internal resistance, R_i, that limits the current under fault conditions. R_i also creates a voltage drop across the barrier. Not accounting for the voltage drop causes the most problems in the proper functioning of intrinsically safe systems.

Examples

Use of barriers for various devices is shown in Figures 15.11(a) through (h).

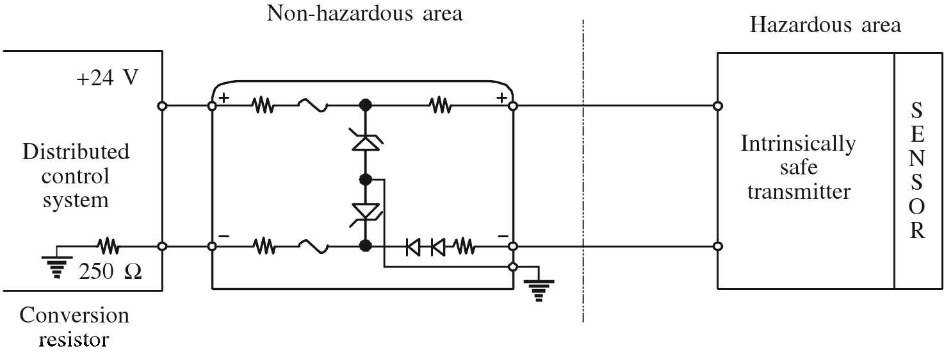


Figure 15.11(a) Supply and return barrier for a 4–20 mA transmitter.

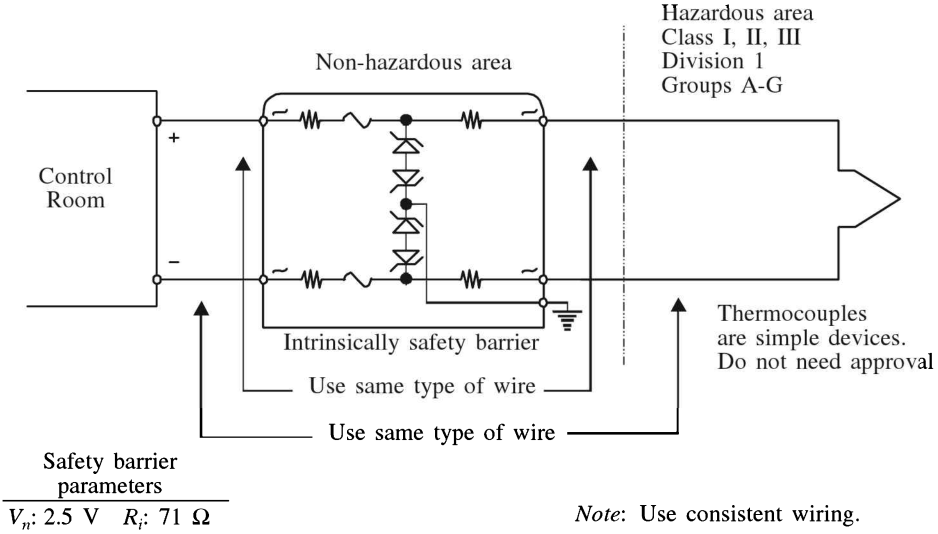
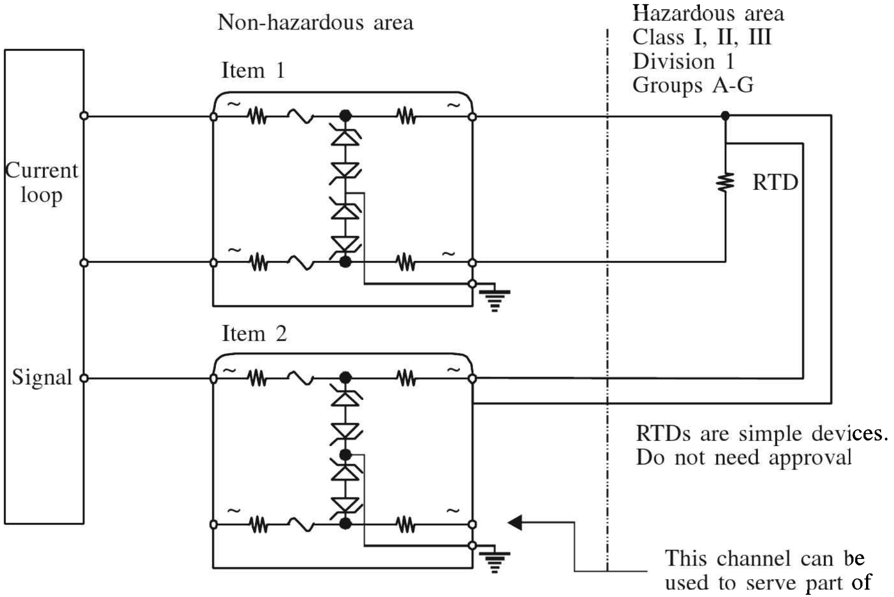


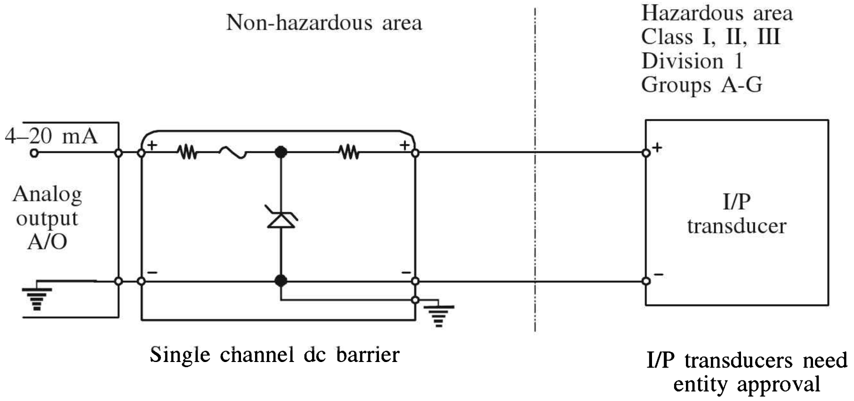
Figure 15.11(b) Typical values of a barrier in thermocouple circuit.



Safety barrier parameters

V_n: 2.5 V R_i: 71 Ω

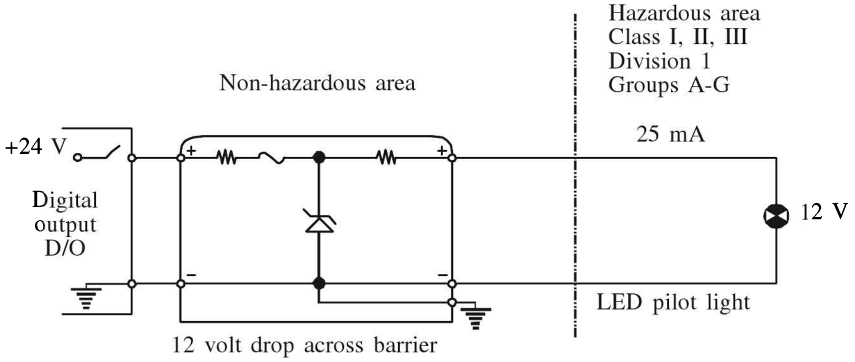
Figure 15.11(c) Intrinsically safe three-wire RTD circuit.



Typical safety barrier parameters

$V_n \geq 12$ V $R_i \geq 150$ Ω

Figure 15.11(d) 4–20 mA I/P transducer.

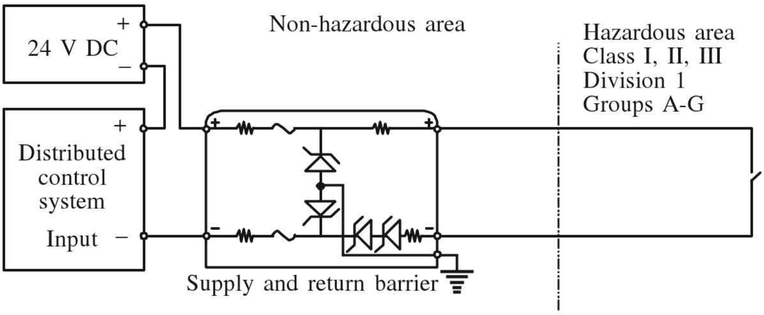


Typical safety barrier parameters

V_n: 24 V R_i: 480 Ω

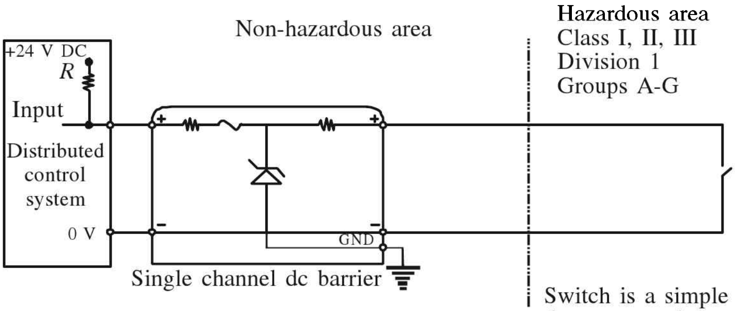
LED pilot light is a simple device; does not need approval

Figure 15.11(e) LED pilot light.



Switch is a simple device; does not need approval

Figure 15.11(f) Current sourcing switching.



Switch is a simple device; does not need approval

Figure 15.11(g) Current sinking switching.

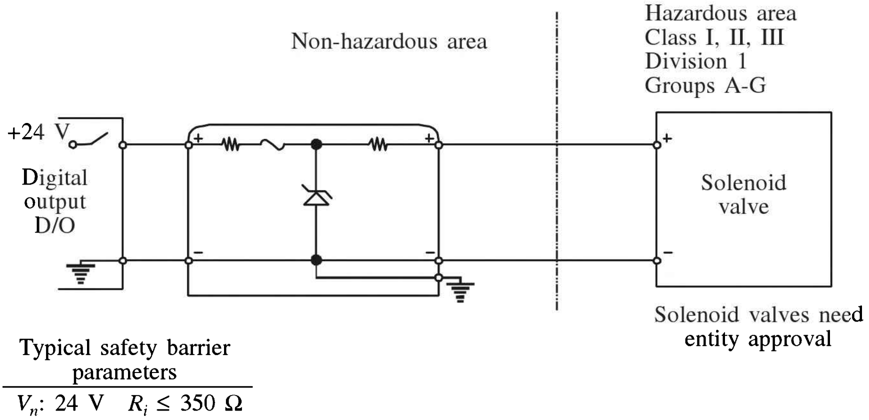


Figure 15.11(h) Connecting a solenoid valve.

Intrinsically safe fieldbus systems

Unlike the 4–20 mA analog loops primarily used in today's process industry, fieldbus networks are digital networks. In a fieldbus network, a repeater ensures transfer accuracy for a digital network. It maintains and enhances the digital signal used to communicate among different fieldbus instruments in both hazardous (explosive) and non-hazardous areas (see Figure 15.12). For a fieldbus isolator, any transfer error is unacceptable. Digital data accuracy

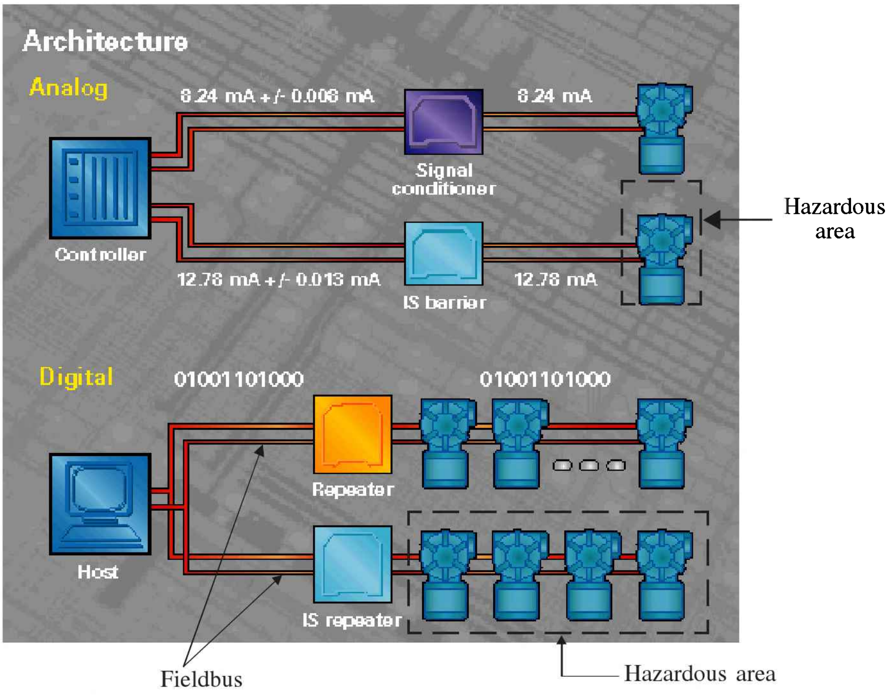


Figure 15.12 IS in analog and digital transmission Hazardous area..

is critical for proper system operation. So while an analog system can tolerate some signal degradation, a digital system cannot. An intrinsic safety barrier for fieldbus must also include a repeater.

A fieldbus repeater provides intrinsically safe isolation/energy limitation, supplies power to the connected field instruments, and refreshes the timing and level of the signal. The latter two functions are crucial for proper implementation of fieldbus systems.

Each network repeater consumes no power on the fieldbus and requires no configuration or addressing. Although it isolates one segment from another, it is transparent to the host and the other connected instruments. Normally, repeaters require independent power from a standard 24-V direct current. Further, the repeater provides an impedance matched power conditioner for the connected fieldbus instruments. Each repeater has internal terminators—necessary impedance placed at the beginning and end of a network segment.

15.4.3 Installing Intrinsically Safe Systems

The intrinsically safe system must be properly installed and provisions must be made to maintain and troubleshoot it. These procedures are discussed in detail in Article 504 of the National Electrical Code (NEC) and the ANSI/ISA RP 12.6-1987 Recommended Practice Installation of Intrinsically Safe Systems for Hazardous (Classified) Locations.

We will study the important techniques and practices used for wiring, installation, grounding, sealing, maintenance, and troubleshooting.

Wiring

Intrinsically safe circuits may be wired in the same manner as comparable circuits installed for unclassified locations with two exceptions summarized as separation and identification. These wiring practices are simple and clear; however, they often are overlooked and are the source of potential problems.

The intrinsically safe conductors must be separated from all other wiring by placing them in separate conduits or by a separation of 2 inches of air space. Within an enclosure the conductors can be separated by a grounded metal or insulated partition (Figure 15.13).

Intrinsically safe wiring colour should be different for identification. The raceways, cable trays, open wiring, and terminal boxes are also labeled Intrinsically Safe Wiring to prevent unintentional interference with the circuits. The spacing between the labels should not exceed 25 ft.

Barrier Installation

The barriers normally are installed in a dust- and moisture-free enclosures located in the non-hazardous area. Only the barrier outputs are intrinsically safe. Conductive dust or moisture could lessen the required distance of 2 inch. between intrinsically safe and nonintrinsically safe conductors (Figure 15.14). The enclosure should be as close as possible to the hazardous area to minimize cable runs and increased capacitance of the circuit. If they are installed in a hazardous area, they must be in the proper enclosure suited for that area.

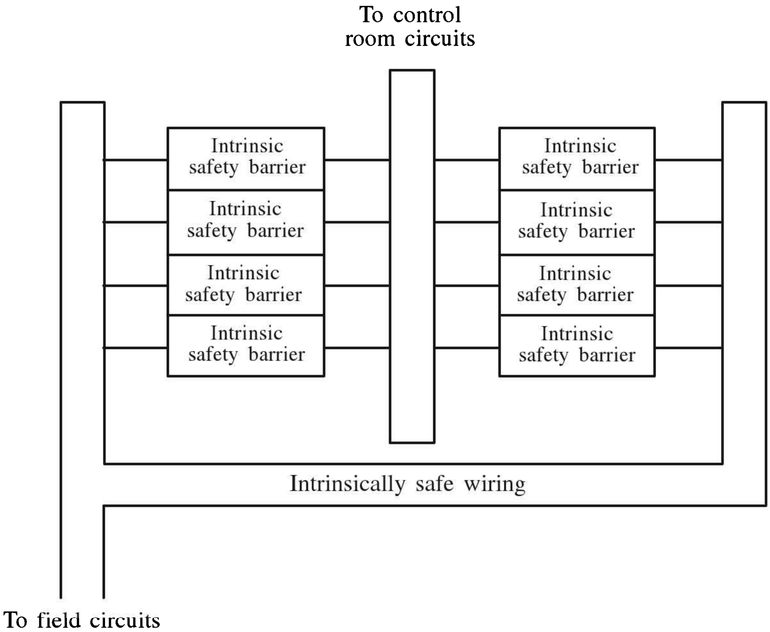


Figure 15.13 Suggested panel arrangement using separate wireways.

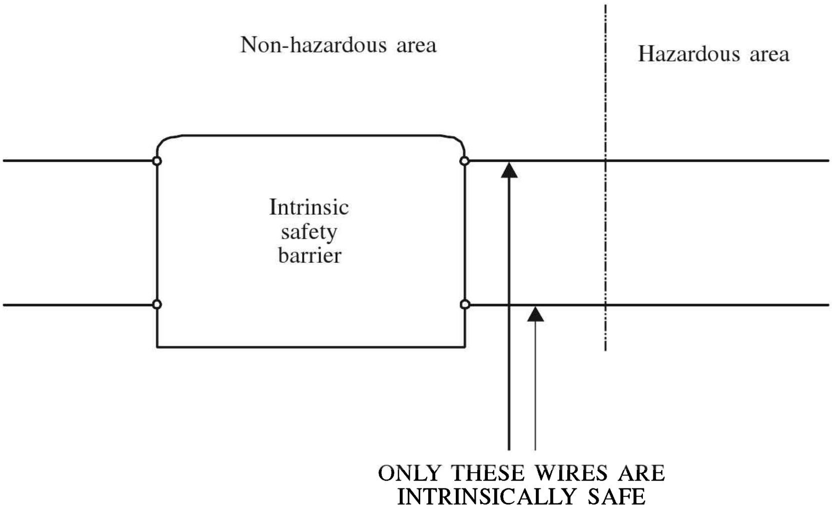


Figure 15.14 Barrier installation.

Grounding

First it is determined if the intrinsically safe barriers used in the system are grounded or isolated. The isolated barriers normally are larger, more expensive, and do not require a ground for safety. The grounded safety barriers are smaller and less expensive, but require a ground to divert the excess energy. The main rules of grounding intrinsically safe systems are:

- The ground path must have less then 1 ohm of resistance from the farthest barrier to the main grounding electrode.
- The grounding conductor must be a minimum 12 AWG.
- All ground path connections must be secure, permanent, visible, and accessible for routine inspection.
- A separate isolated ground conductor normally is required since the normal protective ground conductor (green or yellow/green wire) may not be at the same ground potential because of the voltage drop from fault currents in other equipment.
- For some installations redundant grounding conductors are recommended.

A poor grounding system can influence the function of the system by creating noise on the circuit or modifying the signals. Figure 15.15 shows an improperly grounded system. The numerous grounding points create ground loops which can modify the signals and induce

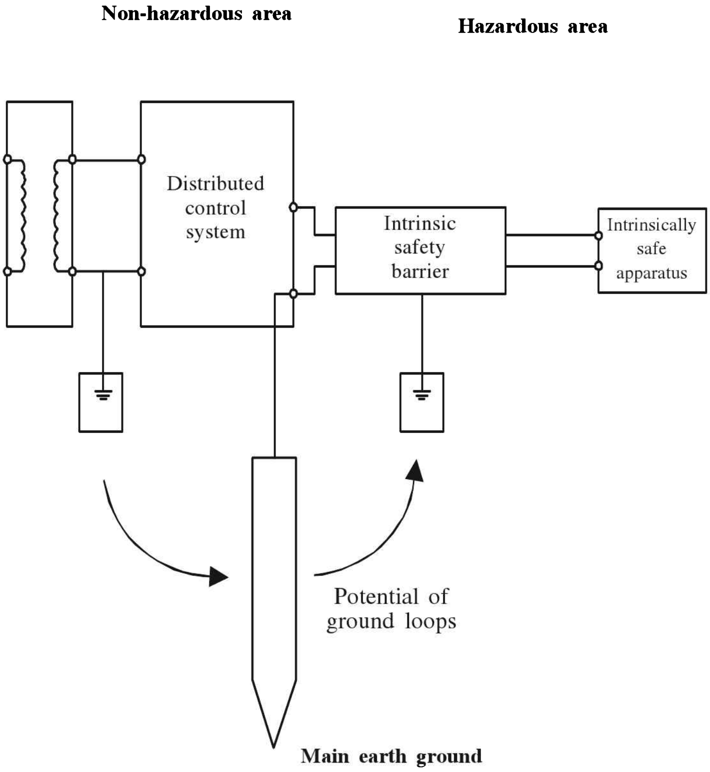

Non-hazardous area
Hazardous area
Distributed control system
Intrinsic safety barrier
Intrinsically safe apparatus
Potential of ground loops
Main earth ground

Figure 15.15 Unacceptable grounding.

stray voltages into the intrinsically safe circuits. The correct method of grounding is shown in Figure 15.16 where all the grounds are tied together at one single point in the system.

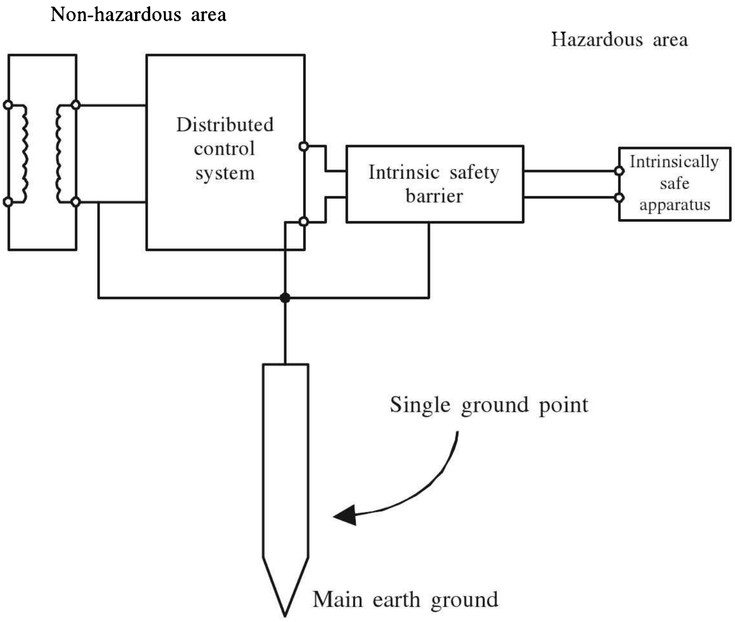


Figure 15.16 Acceptable grounding.

Sealing

Seals are required to prevent the transmission of gases and vapours from the hazardous area to the non-hazardous area, not to prevent passage of flames from explosions. Explosion-proof seals are not required as long as there is some other mechanical means of preventing the passage of gases such as positive pressure in the control room and/or application of an approved mastic at cable terminations and between the cable and raceway. Many experts generally agree that a commercially available silicon caulk is a suitable mastic which would minimize the passage of gases. This must, however, be acceptable to the authority having jurisdiction.

When barriers are installed in explosion-proof enclosures, which are located in the hazardous area, explosion-proof seals are required on the enclosure (Figure 15.17). Since other conduits containing non-intrinsically safe conductors between the hazardous and non-hazardous areas require explosion-proof seals, it is a good practice to maintain consistency and install explosion-proof seals on the conduits containing intrinsically safe conductors as well. The exception to this would be where a multiconductor shielded cable is used. This cable may be difficult to seal in some explosion-proof fittings. However, it is necessary to seal both the cable terminations between the cable and raceway to minimize the passage of gases, vapours, or dust.

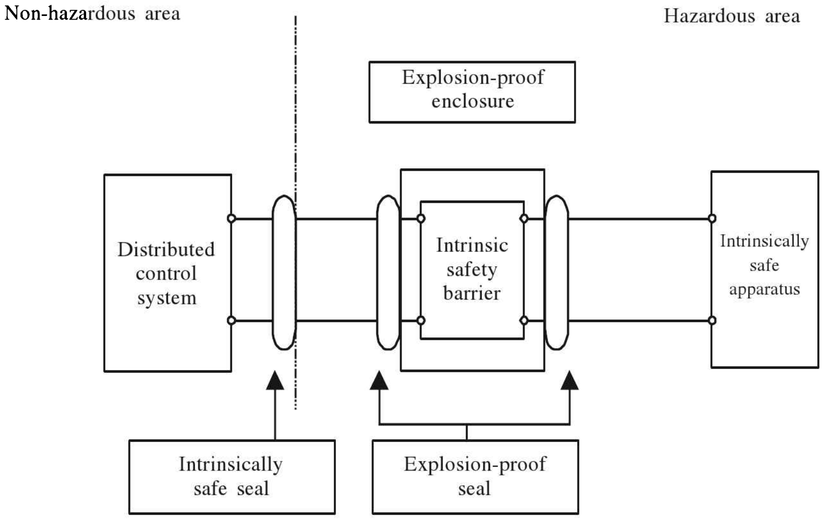


Figure 15.17 Mounting in a hazardous area.

Maintenance

No special maintenance of intrinsically safe systems is required. Once a year the barriers are checked to ensure that the connections are tight, the ground wiring has less than one ohm of resistance, and the barriers are free from moisture and dirt. The panel and conduits are checked for separation and identification of the intrinsically safe wiring. The barrier should not be tested with an ohmmeter or other test instrument while it is connected in the circuit (Figure 15.18). This bypasses the barrier and could induce voltages into the intrinsically safe wiring.

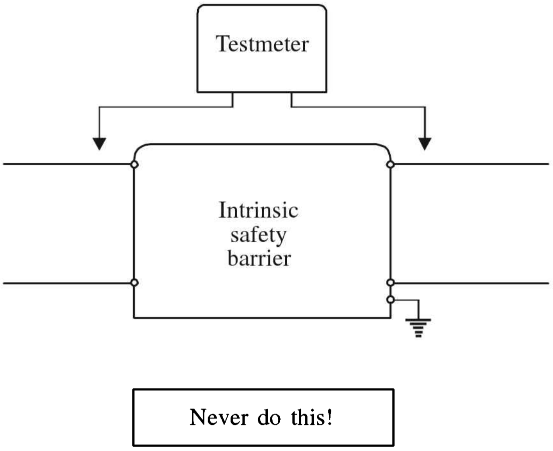


Figure 15.18 The barrier should never be tested with an ohmmeter or other instrument while it is connected in circuit.

Troubleshooting

If the intrinsic safety circuit does not operate properly once it is completed and energized, the following troubleshooting guidelines can be considered:

- Make sure the connections are tight.
- Check the wiring to the appropriate terminals against the control wiring diagram. A control wiring diagram is defined by the NEC as "a drawing or other document provided by the manufacturer of the intrinsically safe or associated apparatus that details the allowed interconnections between the intrinsically safe and associated apparatus." Make sure that one of the manufacturers provides not only diagrams which show the interconnections between the field device and barriers, but also wiring diagrams which demonstrate that the circuit functions properly and is safe by comparing the safety parameters of the field device and the barriers.
- Make sure the circuit is powered.
- Check to see if the resistance in the barrier is too high for the circuit. Circuits are analyzed for the proper loop resistance (barrier and cable) and supply voltages. If the circuit does not operate properly, check the circuit against the design in the control wiring diagram.
- Check for a blown barrier fuse. This is accomplished by disconnecting the barrier from the circuit and measuring the end-to-end resistance of the barrier. If the ohmmeter registers an infinite resistance, the fuse in the barrier is blown. The fuse has opened because of a fault in the circuit, so reevaluate the entire circuit before reinstalling a new barrier.

Barrier replacement

If the barrier's fuse has opened, to replace it, the procedure is to disconnect the wiring from the safety barriers in the proper order of non-hazardous terminal first, hazardous terminals next, and the ground last. Bare wire ends should be covered. Then the barrier should be replaced. To mount the new barrier the procedure should be reversed. Always the ground should be installed first and disconnected the last.

15.4.4 Transformer Isolation Barrier (TIB)

While zener diode barriers provide suitable protection, operate properly and are cost effective, inherent characteristics require close attention to the installation. A high integrity intrinsic safety ground is required to be connected to the ground electrode. There must be no more than one ohm of resistance in this line. A regulated power supply should be utilized. This will prevent the possibility of the voltage rising too high, causing the zener diodes to conduct and the fuse to open, or falling too low preventing the minimum operating voltage from reaching the field device. Process control loops cannot totally float above ground; therefore, it is possible to obtain electrical noise on the control signal. Further, if the protective fuse should open, it cannot be replaced and the barrier must be discarded. Because of these considerations, a different type of barrier known as a Transformer Isolated Barrier (TIB) can be used. It is designed to address each of the disadvantages of zener barriers. A TIB contains a zener barrier for the voltage and current limitation. However, it does not utilize an I.S. ground. The transformer in a TIB provides a high degree of isolation between the primary and secondary windings, therefore, the ground connection is unnecessary.

Advantages of transformer isolated barriers vs. zener diode barriers

(a) Transformer isolated barriers do not require a high integrity I.S. ground since the transformer isolates the hazardous area connections from the non-hazardous area connections.

(b) Regulated power supplies are no longer necessary. A voltage regulator circuit is used to allow a wide supply voltage range without interruption of the process or damage to the barrier.

(c) Any process control loops (milliamp or millivolt signals) connected to a TIB will remain floating, unlike zener barriers that ground one side of the signal or at best provide a quasi-floating system.

(d) TIBs contain current limiting circuitry that will not allow the protective barrier fuse to open in case of a short-circuit condition.

(e) TIBs are designed for specific applications (transmitter, solenoid, switching input, thermocouple, etc.), allowing easy system design.

When to use transformer isolated barriers

(a) When installation of a high integrity I.S. ground is impractical, costly or even impossible.

(b) When regulated power supplies are too expensive.

(c) When the potential of signal noise due to ground loop problems is undesirable.

(d) When the cost of replacing zener barriers due to installation problems (misconceptions and shorted wires) exceeds the initial cost of TIBs.

(e) When the total loop resistance (including barrier, interconnecting cable and safe area load) exceeds the specifications of the transmitter or controller card.

(f) When uncertainty arises during the design of the system due to unfamiliarity with applying barriers.

TIBs are not without a downside. Generally, TIBs cost more than zener barriers and in some cases, need a separate power supply. However, they may be connected in parallel so that one power supply may feed many TIBs. The major objection to using zener diode barriers is that they must be connected to an I.S. ground. This is absolutely necessary to keep a fault in the safe area from reaching the hazardous area. In a TIB, the fault current will flow through the primary of the transformer. Once the current has reached a sufficient level, the fuse will open, therefore preventing the fault condition from reaching the hazardous area.

15.5 RELEVANT INDIAN STANDARDS

A list of relevant Indian standards released by Bureau of Indian Standards is given in Table 15.10.

Table 15.10 List of relevant Bureau of Indian Standards

S.no.	*Indian Standard*	*Subject*
1.	IS 5571:1979/1991	Guide for selection of electrical equipment for hazardous areas
2.	IS 5572:1994	Classification of hazardous areas
3.	IS 7389:1974/1991	Pressurized enclosures of electrical equipment for use in hazardous areas
4.	IS 8240:1976/1991	Guide for electrical equipment for explosive atmosphere
5.	IS 8241:1976/1991	Method of marking for identifying electrical equipment for explosive atmosphere
6.	IS 8945:1987	Electrical measuring instruments for explosive gas atmosphere
7.	IS 9166:1979/1991	Spark test apparatus for intrinsically safe circuits
8.	IS 10398:1982/1991	Zener barrier for intrinsically safe process control instruments
9.	IS 11005:1984/1991	Dust tight ignition proof enclosures of electrical equipment

QUESTIONS

15.1 Explain ignition triangle. What are the sources of ignition? How can ignition be prevented?

15.2 What is the basis of classifications of hazardous areas? Give examples of Class I, II, and III classifications of hazardous areas.

15.3 What is Ex representation scheme? Explain the hazardous area and equipment protection reference scheme.

15.4 (a) What is the range of auto ignition temperature and ignition energy of Class II materials?
(b) What are the two most accepted methods to avoid explosion?

15.5 Explain Ingress Protection (IP) code for the enclosures.

15.6 Explain the protection provided by NEMA enclosures.

15.7 (a) What is the difference between IP code for the enclosures and NEMA enclosures?
(b) What does IP 65 signify? What is its NEMA equivalent?

15.8 (a) Explain pressurized (purged) enclosures used in Class I areas.
(b) What are the cooling arrangements used for cooling electronics in the enclosures? Explain solid state cooling.

15.9 What is the necessity of providing flame paths in enclosures? Give some examples.

15.10 What is the necessity of using seals in enclosures? Give some examples.

15.11 What is meant by intrinsic safety? What is the advantage of using intrinsically safe circuits?

15.12 Draw ignition curves for Group A, B, C, and D areas. Mark safe areas on the curves.

15.13 Explain LOOP and ENTITY approval concepts. What are the conditions of interconnection between hazardous and non-hazardous area products?

15.14 What is meant by intrinsically safe barrier? How does it provide safety? What types of barriers are used? Give typical parameter values of a Class II barrier.

15.15 Is a thermocouple intrinsically safe? If no, how can the circuit of a thermocouple be made safe?

15.16 (a) Give some typical parameters to select a barrier.
(b) Show the intrinsically safe interconnection of the following sensors/devices installed in an hazardous area with the control room in a non-hazardous area:

- 4-20 mA transmitter
- three-wire RTD
- LED pilot lamp
- Solenoid valve
- Relay

15.17 What are the practices used for wiring of intrinsically safe circuits and barrier installation? Suggest a typical panel wiring arrangement.

15.18 What are the considerations for grounding and sealing intrinsically safe systems?

15.19 Give troubleshooting and service guidelines of intrinsically safe circuits.

15.20 What are transformer isolation barriers? Give advantages and disadvantages of transformer isolation barriers over zener diode barriers. When should transformer isolation barriers be used?

Appendices

Appendix A

Attenuators

A.1 INTRODUCTION

An attenuator is used to reduce signal strength by a given amount. Their basic applications include test instruments like multimeters, CRO, signal generators, and signal analyzers. These are also very commonly used between source and amplifiers and between amplifiers and load. The design of attenuators depends on the level of attenuation, characteristic impedance and frequency of operation. Active or passive devices can be used to construct the attenuator. The attenuation can be continuously varying (analog attenuator) or varying in steps (digital or stepped attenuator). We will now discuss simple resistance attenuators.

A.2 ATTENUATION BASICS

In general, attenuation α is defined in decibels. If V_{in} is the input voltage and V_{out} is output voltage of the attenuator, then attenuation is defined as follows:

$$\alpha \text{ in dB} = 20 \log \left(\frac{V_{\text{in}}}{V_{\text{out}}} \right) \tag{A.1}$$

It is assumed that the input and the output resistances are equal.

If we know the desired attenuation in dB, we can calculate $V_{\text{in}}/V_{\text{out}}$ from Eq. (A.1). Let us consider a potential divider as shown in Figure A.1.

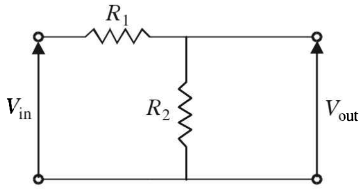


Figure A.1 A potential divider (L-type attenuator).

We have

$$V_{\text{out}} = \left(\frac{R_2}{R_1 + R_2}\right) V_{\text{in}}$$

From Eq. (A.1)

$$\alpha \text{ in dB} = 20 \log \left(1 + \frac{R_1}{R_2}\right) \tag{A.2}$$

If α in dB is given, assuming R_2 equal to output resistance, which is normally given, we can calculate the value of R_1.

Example A.1

Let α = 12 dB and output (load) resistance = 50 ohms. For R_2 = 50 ohms, find R_1.

Solution

From Eq. (A.2), R_1 = 150 ohms

Note that the output resistance of the attenuator matches with the desired output resistance but matching of the input resistance has not been considered. Generally, matching of both input and output resistances is important. Impedance matching is normally used in television, telephone and microwave systems where signal wavelength is comparable to the length of transmission line. If a source of resistance 50 ohms is to be connected to a load of 50 ohms through an attenuator as shown in Figure A.2, attenuator must have resistance (impedance) matching both at the input and at the output. If the resistance seen at the input port of the attenuator is equal to the load resistance, this is called *characteristic* (or *image*) *resistance*. We shall denote it as R_0. Therefore, if in this case we use an attenuator of characteristic resistance of 50 ohms, it will match the resistance both on the source and load side. R_0 can be calculated using Eq. (A.3).

$$R_0 = \sqrt{R_{\text{ins}} R_{\text{ino}}} \tag{A.3}$$

where R_{ins} is the input resistance of the attenuator with the output terminals shorted and R_{ino} is the input resistance of the attenuator with the output terminals opened.

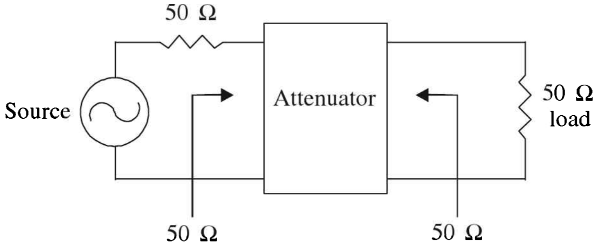


Figure A.2 Impedance matching of attenuator with source and load.

A.3 SYMMETRICAL T ATTENUATOR

Figure A.3 shows a T attenuator. If R_3 = R_1, this is called a symmetrical T attenuator. Since one terminal between input and output is common, this is also called the unbalanced symmetrical T attenuator. In our discussion, we will consider only the unbalanced symmetrical attenuators.

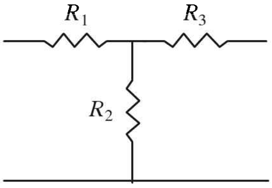


Figure A.3 T attenuator.

Let $R_2 = mR_1$. From Eq. (A.3),

$$R_0 = R_1\sqrt{(1+2m)} \tag{A.4}$$

If we connect a load of value R_0 to this attenuator, then attenuation α will be given by

$$\alpha = \frac{1+m+\sqrt{1+2m}}{m} \tag{A.5}$$

So given R_1 and R_2, we can calculate R_0 and α from Eqs. (A.4) and (A.5). Alternately, an attenuator can be designed for given R_0 and α by the following equations:

$$R_1 = \frac{\alpha-1}{\alpha+1}R_0 \tag{A.6}$$

$$R_2 = \frac{2\alpha}{\alpha^2-1}R_0 \tag{A.7}$$

Example A.2

Find R_0 and α for R_1 = 409 ohms and R_2 = 101 ohms.

Solution

The value of m = 101/409 = 0.247.

From Eqs. (A.4) and (A.5), R_0 = 500 ohms and α = 10 or α in dB = 20 dB.

Example A.3

Design a 20 dB 50 ohms T attenuator.

Solution

Here α = 10 and R_0 = 50 ohms. From Eqs. (A.6) and (A.7), we get

R_1 = 40.9 ohms and R_2 = 10.1 ohms

A.3.1 Cascading T Attenuators

From Eq. (A.7), we observe that as α increases, the value of R_2 decreases and may become impracticably small. Since input and output resistances of these attenuators are same, it is possible to cascade a number of such attenuators to increase the overall attenuation.

Example A.4

Design a 60 dB 500 ohms attenuator using T-section attenuators.

Solution

We can cascade three T-section attenuators, each providing 20 dB attenuation.

From Example A.3, the values of R_1 and R_2 for 20 dB attenuation and 50 ohms characteristic resistance are 40.9 and 10.1 ohms. For the same attenuation with 500 ohms characteristic resistance, R_1 = 409 ohms and R_2 = 101 ohms. A cascaded network with 60 dB attenuation and 500 ohms characteristic resistance is shown in Figure A.4.

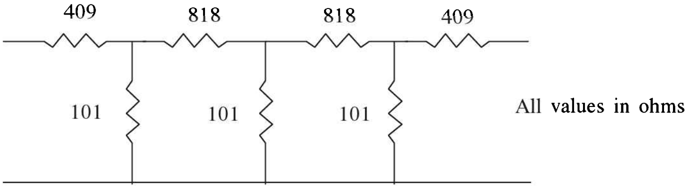


Figure A.4 Cascaded T-section attenuator.

A.4 Symmetrical π Attenuator

Figure A.5 shows an unbalanced symmetrical π attenuator.

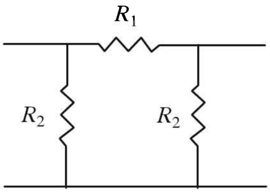


Figure A.5 Symmetrical π attenuator.

Let $R_2 = mR_1$. The equations for calculating R_0 and α are given as

$$R_0 = R_1\sqrt{\frac{m}{1+2m}} \tag{A.8}$$

$$\alpha = \frac{1+m+\sqrt{1+2m}}{m} \tag{A.9}$$

Following design equations can be derived from these equations.

$$R_1 = \frac{\alpha^2 - 1}{2\alpha} R_0 \tag{A.10}$$

$$R_2 = \frac{\alpha+1}{\alpha-1} R_0 \tag{A.11}$$

Cascading of π section networks can be done to achieve higher attenuation levels.

A.5 THE BRIDGED T ATTENUATOR

Figure A.6 shows the Bridged T attenuator.

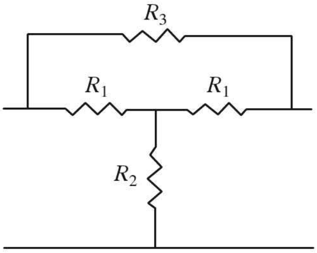


Figure A.6 Bridged T attenuator.

The equations for calculating R_0 and α are given as

$$R_0 = R_1 \tag{A.12}$$

$$\alpha = \frac{R_1}{R_2} + 1 \tag{A.13}$$

It is assumed that $R_1^2 = R_2 \, . \, R_3$.

Following design equations can be derived from these equations.

$$R_1 = R_0 \tag{A.14}$$

$$R_2 = \frac{R_0}{\alpha - 1} \tag{A.15}$$

$$R_3 = (\alpha - 1)R_0 \tag{A.16}$$

Note that R_1 is independent of α.

A.6 VARIABLE ATTENUATORS

In many instruments, variable attenuation levels are needed keeping R_0 constant.

By varying resistors in T or π-section attenuators, we can get continuously varying attenuators. But all the three resistors need to change simultaneously. This may not be an easy proposition. In a bridged T attenuator, we observe from Eqs. (A.14) to (A.16) that R_1 is independent of attenuation, R_2 decreases and R_3 increases with attenuation. Therefore, R_1 can be fixed equal to the value of characteristic resistance. Resistors R_2 and R_3 can be ganged together to vary them simultaneously (Figure A.7).

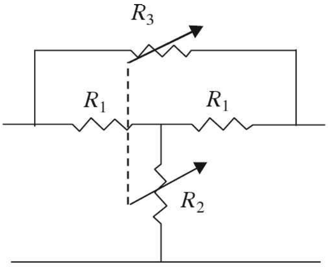


Figure A.7 Continuously variable attenuator.

A.7 STEPPED ATTENUATORS

Another way to achieve variable attenuation is to incorporate various attenuators with different level of attenuation and then select the attenuation through switches. For example, if we construct 1, 2, 3 and 4 dB attenuators, we can get any attenuation level from 1 dB to 10 dB in steps of 1 dB.

Figure A.8 shows two circuits of stepped attenuators. Figure A.8(a) uses a simple resistor network, whereas Figure A.8(b) uses two T attenuators. Given the attenuation desired for each step and impedance requirements, we can design these attenuators using simply the voltage division principle and the formulae derived for the attenuators respectively.

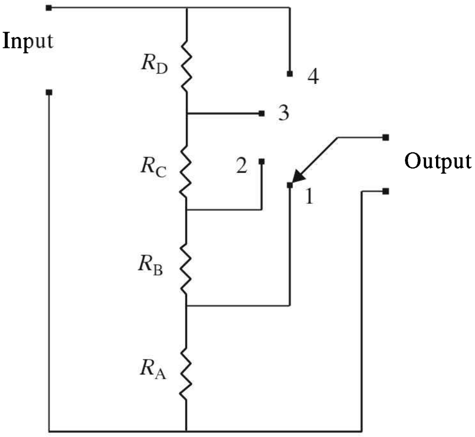
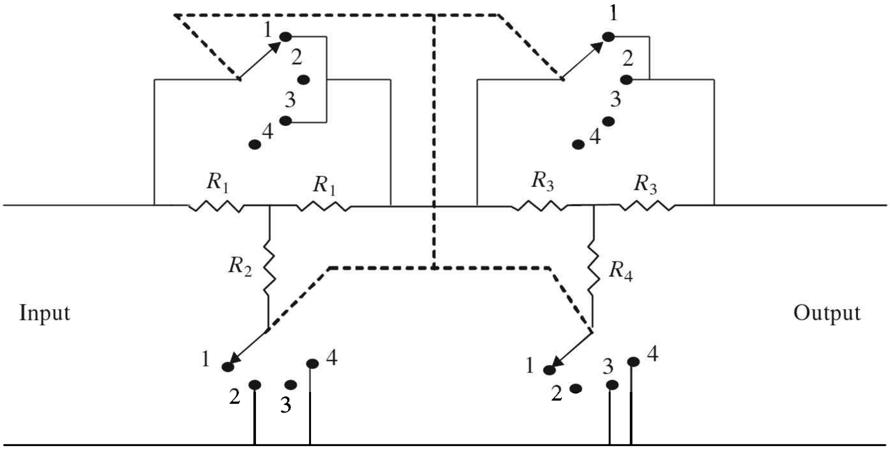


Figure A.8(a) Stepped attenuator: a simple resistor network.



All switches are ganged.

(b)

Figure A.8 Stepped attenuators.

A.8 PADDING

In many applications, the source resistance may not match with the load. Attenuators will see different resistances at the input and output. For example, if source resistance changes with aging or temperature, matching of resistance with load will be affected. It is possible to design an attenuator, to be inserted between source and load, to provide proper matching with the load. Such a use of attenuator is called *padding a source*.

Consider a T attenuator cascaded with a source of resistance R_S as shown in Figure A.9.

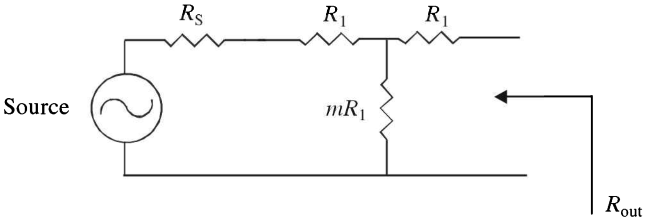


Figure A.9 Padding a source.

Thevenin resistance, R_{out} of this circuit is given by

$$R_{out} = R_1 + mR_1 \parallel (R_1 + R_S) \tag{A.17}$$

Using Eqs. (A.4) and (A.5), Eq. (A.17) can be manipulated as

$$R_{out} = R_0 \left(1 + \frac{2}{\dfrac{\alpha^2 (K+1)}{K-1} - 1} \right) \tag{A.18}$$

where $K = R_S/R_0$. This can be shown that even if π attenuator or any other attenuator is used, Eq. (A.18) is still valid. Note that the larger the value of α, i.e. the larger the attenuation, the more closer is R_{out} to R_0.

Let $R_0 = 50$, $R_S = 150$ and $\alpha = 2$. From Eq. (A.18), $R_{out} = 64.3$ ohms.

Now if for the same values of R_0 and R_S, α is increased to 10, then $R_{out} = 50.5$ ohms which is more closer to R_0. Therefore it is possible to get Thevenin resistance of the source closer to R_0 at the cost of attenuation. In practice, this is acceptable and is used in the design of attenuators.

Variation of R_{out} with R_S

We can find the range of R_{out} for extreme values of R_S for a fixed value of α. Let $R_S = 0$, i.e. $K = 0$. Then

$$R_{out} = R_0\left[1 - \frac{2}{\alpha^2 + 1}\right] \tag{A.19}$$

For $R_S = \infty$, i.e. $K = \infty$. R_{out} is given by

$$R_{out} = R_0\left[1 + \frac{2}{\alpha^2 - 1}\right] \tag{A.20}$$

Let $R_0 = 50$ and $\alpha = 10$. For $R_S = 0$, from Eq. (A.19), $R_{out} = 49$ ohms. For $R_S = \infty$, from Eq. (1.20), $R_{out} = 51$ ohms. Therefore for any value of R_S, R_{out} lies between 49 and 51 ohms. If α is increased further, the difference between R_{out} and R_0 will reduce.

Example A.5

A signal source with an unknown resistance is to be interfaced with a 600 ohms system using a 12 dB 600 ohms pad. Find the extreme values of the padded source resistance.

Solution

Here $\alpha = 4$ and $R_0 = 600$ ohms. From Eqs. (A.19) and (A.20), R_{out} will lie between 528 and 678 ohms.

QUESTIONS

A.1 An oscilloscope is to have a vertical input resistance of 8 MΩ, a sensitivity of 50 mV (maximum input to vertical amplifier) and attenuation factors of 4, 10 and 40. (see Figure QA.1). Compute the values of attenuating resistors R_A, R_B, R_C, and R_D. Ignore input resistance of the vertical amplifier.

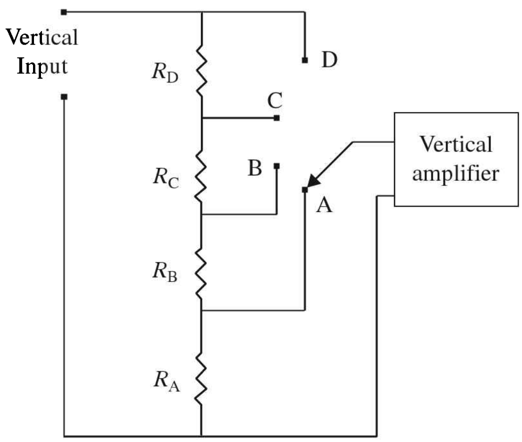


Figure QA.1

A.2 A signal gives minimum output of 100 mV. Its source resistance is 500 Ω. It is desired to feed 1 mV signal to a 500 Ω load. Design a symmetrical T-type attenuator to interface the signal source with load. Do not use resistor values less than 100 Ω.

A.3 Cascaded T-section attenuators, as shown in Figure QA.3, are used between the source with resistance R_S and load R_L. R_S and R_L are made equal to characteristic resistance, R_0. Calculate R_0 and V_{out} if the input voltage to the attenuator is 10 V. Drive the relations used.

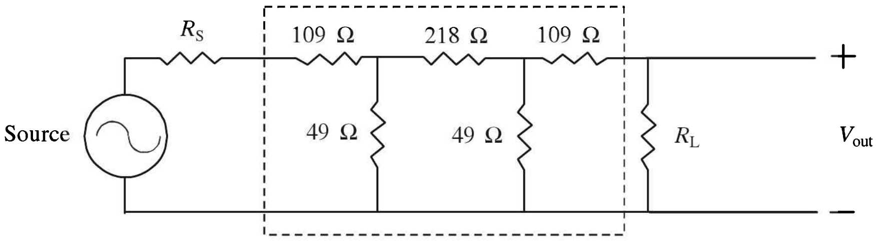


Figure QA.3

A.4 Design a 30 dB 600 Ω π-section attenuator.

A.5 A 100 ohm source is padded by a 6 dB 50 ohm attenuator. What is the Thevenin resistance of the padded source? If the unpadded source resistance changes from 100 to 150 ohms, what is the Thevenin resistance of the padded source?

Appendix B

Terminology Used in Hazardous Area Applications

Adequately Ventilated Area: An adequately ventilated area is an area that has a ventilation system (natural or artificial) that, as a minimum, prevents the accumulation of gases or vapours to an explosive level. Most standards and recommended practices recommend preventing levels in excess of 25 per cent of the Lower Flammable (Explosive) Limit (LFL/LEL). *Note*: Adequate ventilation of an area alone is not an effective way for the prevention of dust explosions.

Approved: Acceptable to the authority having jurisdiction (*see* **also Authority having jurisdiction**). In determining the acceptability of installations or procedures, equipment, or material, the authority having jurisdiction may base acceptance on compliance with appropriate standards. In the absence of such standards, the said authority may require evidence of proper installation, procedure, or use. The authority having jurisdiction may also refer to the listing or labeling practices of product-testing organizations. These organizations are in a position to determine compliance with appropriate standards for the current production of listed or labeled items.

Arcing Apparatus: An electrical make/break component, that is generally interpreted as capable of producing an arc with energy sufficient to cause ignition of a specific ignitable mixture.

Associated Apparatus (*see* also Intrinsic Safety): Apparatus in which the circuits are not intrinsically safe themselves but affect the energy in the intrinsically safe circuits and are relied upon to maintain intrinsic safety. Associated electrical apparatus may be either: (a) electrical apparatus that has an alternative type of protection for use in the appropriate hazardous (classified) location, or (b) electrical apparatus that is not protected and, therefore, cannot be used within a hazardous (classified) location.

Authority having Jurisdiction: The organization, office, or individual that has the responsibility and authority for approving equipment, installations, or procedures. The term *Authority having Jurisdiction* is used in a broad manner since *jurisdiction* and *approval* agencies vary, as do their responsibilities. Where public safety is primary, the authority having jurisdiction may be a central, state, local, or an individual such as an inspector from a labour or health department, electrical inspector, or others having statutory authority. An insurance inspection agency, rating bureau, or other insurance company representative may be the *authority having jurisdiction.* An owner or his designated agent may also assume the role. At government-owned installations, the officer-in-charge, departmental official, or designated agent may be the *authority having jurisdiction.*

Bonding: The permanent joining of metallic parts to form an electrically conductive path that will ensure electrical continuity and the capacity to conduct safely any current likely to be imposed.

Certified: Generic term referring to equipment that has been evaluated by a recognized testing agency and confirmed to be in compliance with the applicable standard(s). Some agencies use the terms *approved*, *listed*, or *labeled equipment* to indicate compliance with the acceptable standard.

Class I Location: A location in which flammable gases or vapours are or may be present in the air in quantities sufficient to produce explosive or ignitable mixtures.

Class I, Division 1 Location: A location (a) in which ignitable concentrations of flammable gases or vapours can exist under normal operating conditions, (b) in which ignitable concentrations of such gases or vapours may exist frequently because of repair or maintenance operations or because of leakage, or (c) in which breakdown or faulty operation of equipment or processes might release ignitable concentrations of flammable gases or vapours and might also cause simultaneous failure of electrical equipment that could act as a source of ignition.

Class I, Division 2 Location: A location (a) in which volatile flammable liquids or flammable gases are handled, processed, or used, but in which the liquids, vapours, or gases will normally be confined within closed containers or closed systems from which they can escape only in case of accidental rupture or breakdown of such containers or systems, or in case of abnormal operation of equipment, or (b) in which ignitable concentrations of gases or vapours are normally prevented by positive mechanical ventilation and might become hazardous through failure or abnormal operation of the ventilating equipment, or (c) that is adjacent to a Class I, Division 1 location and to which ignitable concentrations of gases or vapours might occasionally be communicated unless such communication is prevented by adequate positive-pressure ventilation from a source of clean air and effective safeguards against ventilation failure are provided.

Class II Location: A location that is hazardous because of the presence of combustible dust.

Class II, Division 1 Location: A location (a) in which combustible dust is in the air under normal operating conditions in quantities sufficient to produce explosive or ignitable mixtures, or (b) in which mechanical failure or abnormal operation of machinery or equipment might cause such explosive or ignitable mixtures to be produced and might also provide a source of ignition through simultaneous failure of electrical equipment, operation of protection devices, or from other causes, or (c) in which combustible dusts of an electrically conductive nature may be present in hazardous quantities.

Class II, Division 2 Location: A location in which combustible dust is not normally in the air in quantities sufficient to produce explosive or ignitable mixtures, and dust accumulations are normally insufficient to interfere with the normal operation of electrical equipment or other apparatus, but combustible dust may be in suspension in the air as a result of infrequent malfunctioning of handling or processing equipment and where combustible dust accumulations on, in, or in the vicinity of the electrical equipment may be sufficient to interfere with the safe dissipation of heat from electrical equipment or may be ignitable by abnormal operation or failure of electrical equipment.

Class III Location: A location that is hazardous because of the presence of easily ignitable fibers or flyings but in which such fibers or flyings are not likely to be in suspension in the air in quantities sufficient to produce ignitable mixtures.

Class III, Division 1 Location: A location in which easily ignitable fibers or materials producing combustible flyings are handled, manufactured, or used.

Class III, Division 2 Location: A location in which easily ignitable fibers are stored or handled (except in the process of manufacture).

Control Drawing: A drawing or other document provided by the manufacturer of the intrinsically safe or associated apparatus that details the allowed interconnections between the intrinsically safe and associated apparatus.

Degree of Protection (IP) (*see* also Enclosure Type): A system of rating standard levels of protection provided by apparatus for the protection of persons against contact with live or moving parts inside the apparatus, as well as the protection provided by apparatus against ingress of solids and/or liquids. This type of protection classification is in addition to (and not an alternative to) the types of protection necessary to ensure protection against ignition in hazardous (classified) locations. Definitions are found in IEC Publication 529.

Dust, Combustible: Any finely divided solid material 420 microns or less in diameter that presents a fire or explosion hazard.

Dust-Ignition Proof: A term used to describe an enclosure that will exclude ignitable amounts of dust that might affect performance or rating and that, when installed in accordance with the original design intent, will not permit arcs, sparks, or heat otherwise generated or liberated inside the enclosure to cause ignition of exterior accumulations or atmosphere suspensions of a specified dust in the vicinity of the enclosure.

Dust Layer, Combustible: Any surface accumulation of combustible dust that is thick enough to propagate flame or will degrade and ignite.

Dust-Protected Enclosure: A term describing an enclosure in which the ingress of dust is not totally prevented, but does not enter in sufficient quantity to interfere with the safe operation of the equipment or accumulate in a position within the enclosure where it is possible to cause an ignition hazard.

Dust Tight: An enclosure so constructed that dust will not enter the enclosing case under specified test conditions.

EEx: Designation of explosion protected apparatus complying with European Standards.

Ex: Designation of explosion protected apparatus.

Ex Scheme: An international system of certification for explosion protected apparatus administered by the IEC and described by IECEE-04. The eventual goal of this scheme is that a manufacturer of hazardous location electrical equipment would be able to obtain a single 'Ex' Certificate of Conformity from one Certification Laboratory and sell that product in any participating country without legal or technical obstacles and without the need to get it recertified locally.

Encapsulation: A type of protection in which the parts that could ignite an explosive atmosphere by either sparking or heating are enclosed in an encapsulant in such a way that this explosive atmosphere cannot be ignited. This type of protection is referred to as ‘m’.

Enclosure Type (*see* also Degree of Protection): A system of rating standard levels of protection provided to electrical apparatus by enclosures for (a) the protection of persons against contact with live or moving parts inside the enclosure, (b) the protection provided by enclosure against ingress of solids and/or liquids, (c) the protection provided by the enclosure against the deleterious effects of corrosion, and (d) the protection provided by the enclosure against damage due to the formation of external ice. This enclosure type is in addition to (and not an alternative to) the types of protection necessary to ensure protection against ignition in hazardous (classified) locations.

Entity Concept: A concept that allows interconnection of intrinsically safe apparatus to associated apparatus not specifically examined in such a combination. The criteria for interconnection is that the voltage (V_{max}) and current (I_{max}) which intrinsically safe apparatus can receive and remain intrinsically safe, considering faults, must be equal to or greater than the voltage (V_{oc} or V_t) and current (I_{sc} or I_t) levels which can be delivered by the associated apparatus, considering faults and applicable factors. In addition, the maximum unprotected capacitance (C_i) and inductance (L_i) of the intrinsically safe apparatus, including interconnecting wiring, must be equal to or less than the capacitance (C_a) and inductance (L_a) that can safely be connected to the associated apparatus. If these criteria are met, then the combination may be connected.

Entry, Direct: A method of connection of an electrical apparatus to the external circuits by means of the connecting facilities inside the main enclosure or in a terminal compartment having a free opening to the main enclosure.

Entry, Indirect: A method of connection of an electrical apparatus to the electrical circuits by means of a terminal box or a plug and socket connection which is external to the main enclosure.

Explosion-proof Apparatus (*see* also Flameproof Enclosure): Apparatus enclosed in a case that is capable of withstanding an explosion of a specified gas or vapour that may occur within it and of preventing the ignition of a specified gas or vapour surrounding the enclosure by sparks, flashes, or explosion of the gas or vapour within, and that operates at such an external temperature that a surrounding flammable atmosphere will not be ignited thereby.

Explosive Atmosphere (*see* also Hazardous (classification) Location): A mixture with air, under atmospheric conditions, of flammable substances in the form of gas, vapour, mist, dust, or fibers in which, after ignition, combustion spreads throughout the unconsumed mixture.

Fault (As Applicable to Intrinsically Safe Systems) (*see* also Protective Component): A defect or electrical breakdown of any component, spacing, or insulation that alone or in combination with other defects or breakdowns may adversely affect the electrical or thermal characteristics of the intrinsically safe system. If a defect or breakdown leads to defects or breakdowns in other components, the primary and subsequent defects and breakdowns are considered to be single fault. Certain components may be considered not subject to fault when analyses or tests for intrinsic safety are made.

Fibers and Flyings: These are materials not normally in suspension in air; and are of larger particle size than dusts. Fibers and flyings include materials such as cotton linters, sawdust, textile fibers, and other particles that are usually more a fire hazard than an explosion hazard.

Flameproof Enclosure: A type of protection of electrical apparatus in which the enclosure will withstand an internal explosion of a flammable mixture which has penetrated into the interior, without suffering damage and without causing ignition, through any joints or structural openings in the enclosure, of an external explosive atmosphere consisting of one or more of the gases or vapours for which it is designed. This type of protection is referred to as 'd'.

Flammable (Explosive) Limits: The flammable (explosive) limits of a gas or vapour are the lower (LFL/LEL) and upper (UFL/UEL) flammable (explosive) limit percentages by volume of concentration of gas in a gas-air mixture that will form an ignitable mixture.

Flammable Liquid: Any liquid having a flash point below 37.8°C (100°F) and having a vapour pressure not exceeding 275 kPa (40 psia) at 37.8°C (100°F).

Flammable Gas or Vapour: A gas or vapour which, when mixed with air in certain proportions, will form an explosive gas atmosphere.

Flash Point: The minimum temperature at which a liquid gives off vapour in sufficient concentration to form an ignitable mixture with air near the surface of the liquid, as specified by test.

Group: A classification of electrical apparatus related to the explosive atmosphere for which it is to be used.

Hazardous (Classification) Location (*see* also Explosive Atmosphere): A location in which fire or explosion hazards may exist due to an explosive atmosphere of flammable gases or vapours, flammable liquids, combustible dust, or easily ignitable fibers or flyings.

Hermetically Sealed Device: A device that is sealed against the entrance of an external atmosphere and in which the seal is made by fusion, e.g. soldering, brazing, welding, or the fusion of glass to metal.

High Temperature Apparatus: As apparatus in which the maximum operating temperature (including ambient temperature effect) exceeds 80 per cent of the autoignition temperature in degrees Celsius (°C) of the gas or vapour involved.

Ignition (Autoignition) Temperature (AIT) (*see* also Flash Point): The minimum temperature required to initiate or cause self-sustained combustion of a solid, liquid, or gas independently of the heating or heating elements. A distinction is made between ignition temperature and flash point.

Ignition Capable: Equipment or wiring that under normal conditions, or under specified abnormal conditions, can release sufficient electrical or thermal energy to cause ignition of a specific explosive atmosphere.

Increased Safety: Type of protection applied to electrical apparatus that does not produce arcs or sparks in normal service and under specified abnormal conditions, in which additional measures are applied so as to give increased security against the possibility of excessive temperatures and of the occurrence of arcs and sparks. This type of protection is referred to as 'e'.

Intrinsic Safety (*see* also Associated Apparatus): A type of protection in which a portion of the electrical system contains only intrinsically safe equipment (apparatus, circuits and wiring) that is incapable of causing ignition in the surrounding atmosphere. No single device or wiring is intrinsically safe by itself (except for battery-operated self-contained apparatus such as portable pagers, transceivers, gas detectors, etc. which are specifically designed as intrinsically safe self-contained devices) but is intrinsically safe only when employed in properly designed intrinsically safe system. This type of protection is referred to as 'i'.

Intrinsical Safety Barrier: A component containing a network designed to limit the energy (voltage and current) available to the protected circuit in the hazardous (classified) location under specified fault conditions.

Intrinsically Safe Circuit: A circuit in which any spark or thermal effect, produced either normally or in specified fault conditions, is capable, in the specified test conditions, of causing ignition of a given explosive atmosphere.

Intrinsically Safe Electrical Apparatus: Electrical apparatus in which all the circuits are intrinsically-safe circuits.

Intrinsical Safety Ground Bus: A grounding system that has a dedicated conductor separate from the power system so that the ground currents will not normally flow and that is reliably connected to a ground electrode.

Intrinsically Safe System: An assembly of interconnected intrinsically safe apparatus, associated apparatus, other apparatus, and interconnecting cables in which those parts of the system that may be used in hazardous (classified) locations are intrinsically-safe circuits.

Liquid, Combustible: A liquid having a flash point at or above 37.8°C (100°F). Combustible liquids are subdivided as follows: (a) Class II liquids include those having flash points at or above 37.8°C (100°F) and below 60°C (140°F); (b) Class IIIA liquids include those having flash points at or above 60°C (140°F) and below 93°C (200°F); (c) Class IIIB liquids include those having flash points at or above 93°C (200°F). It should also be noted that these 'Classes' have no relation to the hazardous location 'Classes'.

Listed: Equipment or materials included in a list published by an organization acceptable to the Authority having jurisdiction and concerned with product evaluation, that maintains periodic inspection of production of listed equipment or materials, and whose listing states either that the equipment or material meets appropriate designated standards or has been tested and found suitable for use in a specified manner. The means for identifying listed equipment may vary for each organization concerned with product evaluation, some of which do not recognize equipment as listed unless it is also labeled. The Authority having Jurisdiction should utilize the system employed by the listing organization to identify a listed product.

Lower Explosive Limit (LEL): Refer to Flammable (Explosive) Limits.

Lower Flammable Limit (LFL): Refer to Flammable (Explosive) Limits.

Make/Break Component: Components having contacts that can interrupt a circuit (even if the interruption is transient in nature). Examples of make/break components are relays, circuit breakers, servo potentiometers, adjustable resistors, switches, connectors, and motor brushes.

Maximum Surface Temperature: The highest temperature attained by a surface accessible to flammable gases, vapours or combustible dusts under conditions of operation within the ratings of the apparatus (including specified abnormal conditions).

Minimum Cloud Ignition Temperature: The minimum temperature at which a combustible dust atmosphere will autoignite and propagate an explosion.

Minimum Dust Layer Ignition Temperature: The minimum temperature of a surface that will ignite a dust on it after a long time (theoretically, until infinity). In most dusts, free moisture has been vaporized before ignition.

Minimum Explosive (Dust) Concentration: The minimum concentration of a dust cloud which, when ignited, will propagate a flame away from the source of ignition.

Minimum Ignition Energy (MIE): The smallest amount of energy that can ignite the most easily ignitable mixture of a specific gas or vapour-in-air mixture or dust-in-air mixture.

Maximum Experimental Safe Gap (MESG): The maximum clearance between two parallel metal surfaces which has been found, under specified test conditions, to prevent an explosion in a test chamber from being propagated to a secondary chamber containing the same gas or vapour at the same concentration.

Minimum Igniting Current Ratio (MIC Ratio): The ratio derived by dividing the minimum current required from an inductive spark discharge to ignite the most easily ignitable mixture of a gas or vapour by the minimum current required from an inductive spark discharge to ignite methane under the same test conditions.

Nonhazardous (Unclassified) Location: A location in which fire or explosion hazards are not expected to exist specifically due to the presence of flammable gases or vapours, flammable liquids, combustible dusts, or ignitable fibers or flyings. Such a location may also be referred to as a Safe Area.

Nonincendive Circuit (*see* also Nonincendive Field Wiring): A circuit, other than field wiring, in which any arc or thermal effect produced, under intended operating conditions of the equipment, is not capable, under specified test conditions, of igniting the flammable gas-, vapour, or dust-in-air mixture.

Nonincendive Component: A component having contacts for making or breaking an ignition-capable circuit and in which the contacting mechanism is constructed so that the component is incapable of igniting the specified explosive atmosphere. The housing of a nonincendive component is not intended to (a) exclude the flammable atmosphere, or (b) contain an explosion. This type of protection is referred to as ‘nA’.

Nonincendive Equipment: Equipment having electrical/electronic circuitry and components that are incapable, under normal conditions, of causing ignition of the flammable gas-, vapour-, or dust-in-air mixture due to arcing or thermal effect. This type of protection is referred to as ‘NA’, ‘NC’, or ‘NR’.

Nonincendive Field Wiring (*see* also Nonincendive Circuit): Wiring that enters or leaves an equipment enclosure and, under normal operating conditions of the equipment, is not capable, due to arcing or thermal effects, of igniting the flammable gas-, vapour-, or dust-in-air mixture. Normal operation includes opening, shorting, or grounding the field wiring.

Normal Conditions: Equipment is generally considered to be under normal conditions when it conforms electrically and mechanically with its design specifications and is used within the limits specified by the manufacturer.

Oil-Immersion: Type of protection in which the electrical apparatus or parts of the electrical apparatus are immersed in a protective liquid in such a way that an explosive atmosphere which may be above the liquid or outside the enclosure cannot be ignited. This type of protection is referred to as ‘o’.

Powder Filling: A type of protection in which the parts capable of igniting an explosive atmosphere are fixed in position and completely surrounded by filling material to prevent the ignition of an external explosive atmosphere. This type of protection is referred to as 'q'. This type of protection may not prevent the surrounding explosive atmosphere from penetrating into the apparatus and Ex components and being ignited by the circuits. However, due to the small free volumes in the filling material and due to the quenching of a flame which may propagate through the paths in the fillings material, an external explosion is prevented.

Pressurization: The technique of guarding against the ingress of the external atmosphere into an enclosure by maintaining a Protective Gas therein at a pressure above that of the external atmosphere. This type of protection is referred to as ‘p’.

Pressurization, Type X: A method of reducing the classification within an enclosure from Division 1/Zone 1 to nonhazardous (unclassified).

Pressurization, Type Y: A method of reducing the classification within an enclosure from Division 1/Zone 1 to Division 2/Zone 2.

Pressurization, Type Z: A method of reducing the classification within an enclosure from Division 2/Zone 2 to nonhazardous (unclassified).

Protection, Type of: The specific measures applied to electrical apparatus to avoid ignition of a surrounding explosive atmosphere. Examples are 'e' and 'n'.

Protective Component (As Applied to Intrinsic Safety): A component that is so unlikely to become defective in a manner that will lower the intrinsic safety of the circuit that it may be considered not subject to fault when analyses or tests for intrinsic safety are made.

Protective Gas: The gas used for pressurization or for the dilution of flammable gases to a level well below their lower explosive limit, usually below 25 per cent LFL/LEL. The protective gas may be air, nitrogen, other nonflammable gas, or a mixture of such gases.

Purging: In a pressurized enclosure, the operation of passing a quantity of protective gas through the enclosure and ducts, so that the concentration of the explosive gas atmosphere is brought to a safe level.

Restricted Breathing: A protection technique in which the tightness of an enclosure is assured so that short-term presence of a flammable gas or vapour cloud around the enclosure will not cause the concentration inside the enclosure to reach the LFL/LEL because of breathing or diffusion. This type of protection is referred to as 'nR'.

Safe Area: Refer to Nonhazardous (Unclassified) Location.

Seal, Cable, Explosion-proof: A cable termination fitting filled with compound and designed to contain an explosion in the enclosure to which it is attached or to minimize passage of flammable gases or vapours from one location to another. A conduit seal may also be used as a cable seal.

Seal, Conduit, Explosion-proof: A sealing fitting, filled with a poured potting compound, designed to contain an explosion in the enclosure to which it is attached and to minimize passage of flammable gases or vapours from one location to another.

Seal, Factory: A construction where components capable of initiating an internal explosion due to arcing, sparking, or thermal effects under normal conditions are isolated from the wiring system by means of factory-installed flameproof seal or joint for the purpose of eliminating the need for an external, field-installed conduit seal and, in some cases, a field-installed cable seal.

Sealed Device: A device so constructed that it cannot be opened during normal operational conditions or operational maintenance; it has a free internal volume less than 100 cubic centimetres (6.1 cubic inches) and is sealed to restrict entry of an external atmosphere. This type of protection is referred to as 'nC'.

Simple Apparatus (As Applied to Intrinsic Safety): A device that will not generate or store more than 1.2 V, 0.1 A, 25 mW, or 20 J. Examples are: switches, thermocouples, light-emitting diodes, and resistance temperature detectors (RTDs).

Source of Release: A point from which flammable gases or vapours, flammable liquid, combustible dusts, or ignitable fibers or flyings may be released into the atmosphere.

Temperature, Ambient: The temperature of air or other media where electrical apparatus is to be used.

Temperature Identification Number (T Code/Temperature Class): A system of classification by which one of 14 temperature identification numbers (for zones, six temperature classes) is allocated to apparatus. The temperature identification number represents the maximum surface temperature of any part of the apparatus that may come in contact with the flammable gas or vapour mixture.

Type of Protection: Refer to Protection, Type of.

Upper Explosive Limit (UEL): Refer to Flammable (Explosive) Limits.

Zone: A method of specifying the probability that a location is made hazardous by the presence, or potential presence, of flammable concentrations of gases and vapours, or combustible mixtures of dusts.

Zone 0 (IEC): An area in which an explosive gas atmosphere is present continuously or for long periods.

Zone 0 (NEC): A Class I, Zone 0 location is a location (a) in which ignitable concentrations of flammable gases or vapours are present continuously or (b) in which ignitable concentrations of flammable gases or vapours are present for long periods of time.

Zone 1 (IEC): An area in which an explosive gas atmosphere is likely to occur in normal operation.

Zone 1 (NEC): A Class I, Zone 1 location is a location (a) in which ignitable concentrations of flammable gases or vapours are likely to exist under normal operating conditions or (b) in which ignitable concentrations of flammable gases or vapours may exist frequently because of repair or maintenance operations or because of leakage, or (c) in which equipment is operated or processes are carried on, of such a nature that equipment breakdown or faulty operations could result in the release of ignitable concentrations of flammable gases or vapours and also cause simultaneous failure of electrical equipment in a mode to cause the electrical equipment to become a source of ignition, or (d) that is adjacent to a Class I, Zone 0 location from which ignitable concentrations of vapours could be communicated, unless communication is prevented by adequate positive-pressure ventilation from a source of clean air and effective safeguards against ventilation failure are provided.

Zone 2 (IEC): An area in which an explosive gas atmosphere is not likely to occur in normal operation and, if it does occur, it is likely to occur only infrequently for a short period only.

Zone 2 (NEC): A Class I, Zone 2 location is a location (a) in which ignitable concentrations of flammable gases or vapours are not likely to occur in normal operation and if they do occur, they will exist only for a short period, or (b) in which volatile flammable liquids, flammable gases, or flammable vapours are handled, processed, or used, but in which the liquids, gases, or vapours normally are confined within closed containers or closed systems from which they can escape only as a result of accidental rupture or breakdown of the containers or system, or as the result of abnormal operation of the equipment with which the liquids or gases are handled, processed, or used, or (c) in which ignitable concentrations of flammable gases or vapours normally are prevented by positive mechanical ventilation, but which may become hazardous as the result of failure or abnormal operation of the ventilation equipment, or (d) that is adjacent to a Class I, Zone 1 location, from which ignitable concentrations of flammable gases or vapours could be communicated, unless such communication is prevented by adequate positive-pressure ventilation from a source of clean air, and effective safeguards against ventilation failure are provided.

Zone 21 (IEC): An area not classified as Zone 20 in which combustible dust, as a cloud, is likely to occur during normal operation, in sufficient quantity to be capable of producing an explosive concentration of combustible dust with air. This zone can include, among others, areas in the immediate vicinity of powder filling or emptying points, and areas where dust layers occur and are likely in normal operation to give rise to an explosive concentration of combustible dust in mixture with air.

Zone 22 (IEC): An area not classified as Zone 21 in which combustible dust, as a cloud, can occur infrequently, and persist only for a short period, or in which accumulations or layers of combustible dust can give rise to an explosive concentration of combustible dust in mixture with air. This zone can include, among others, areas in the vicinity of equipment containing dust, which dust can escape from leaks and form deposits (e.g. milling rooms in which dust can escape from the mills and then settle).

Appendix C

Instrumentation Symbols and P & I Diagrams

C.1 INTRODUCTION

In an industry, the process can be described by the use of instrumentation Piping & Instrumentation (P&I) diagrams. Its purpose is to provide information quickly and systematically for use in process analysis, production control, specification of equipment, and preparation of equipment requisitions.

An instrumentation diagram should indicate the following:

- The variable being measured
- Whether indication, recording or other service is required
- Whether control or alarm functions are required
- The auxiliary features of the instrument or controller
- The type of connecting lines
- The approximate location of the point of measurement and control
- Type of instrumentation to be included in the control centre and at the processing end

To represent these, we use certain symbols and notations. These are based on the Standards (ISA S5.1) prescribed by the Instrumentation Society of America.

C.2 DEFINITIONS

Alarm: A device that signals the existence of an abnormal condition by means of an audible or visible or both discrete change, intended to attract attention.

Balloon: The circular symbol used to denote an instrument or instrument tagging.

Behind the board: A term applied to a location that is

- within an area that contains the instrument board
- within or in the back of the board, or is otherwise not accessible to the operator for his normal use, and
- not designated as local.

Board: A structure that has a group of instruments mounted on it and that is chosen to have an individual designation. The board may consist of one or more panels, cubicles, desks, or racks.

Board Mounted: A term applied to an instrument that is mounted on a board and that is accessible to the operator for his normal use.

Computing Relay: A relay that performs one or more calculations or logical functions or both, and sends out one or more resultant output signals.

Controller: A device that has an output that can be varied to maintain a controlled variable at a specified value or within specified limits or to alter the variable in a specified manner.

An automatic controller varies its output automatically in response to direct or indirect input of a measured input variable. A manual controller is a manual loading station, and its output is not dependent on a measured process variable but can be varied only by manual adjustment. A controller can be integral with other functional elements of a control loop.

Control Station: A manual loading station that also provides switching between manual and automatic control modes of a control loop. It is also known as auto-manual station and an auto-selector station.

Control Valve: A device, other than a common hand actuated on-off valve, that directly manipulates the flow of one or more fluid process streams. In some applications, it is commonly known as a damper or louver.

The designation hand-control valve is limited to hand-actuated valves that (a) are used for process throttling, or (b) are special valves for control purposes and that are to be specified by an instrumentation group or an instrumentation engineer.

Converter: A device that receives information in the form of an instrument signal, alters the form of information, and sends out a resultant output signal. A converter is a special form of relay. A converter is also referred to as a transducer, although transducer is a completely general term and its use specifically for signal conversion is not recommended.

Final Control Element: The device that directly changes the value of the manipulated variable of a control loop.

Function: The purpose of or action performed by a device.

Identification: The sequence of letters or digits or both used to designate an individual instrument or loop.

Instrument: A device used directly or indirectly to measure or control a variable. The term includes control valves, relief valves, and electrical devices such as annunciators and push buttons. This term does not apply to parts, e.g. a receiver bellow or a resistor, that are internal components of an instrument.

Instrumentation: The application of instruments.

Local: The location of an instrument that is neither on or behind the board. Local instruments are commonly used in the vicinity of a primary element or a final control element.

Local Board: A board that is not a central or main board. Local boards are commonly used in the vicinity of plant subsystems or sub-areas.

Loop: A combination of one or more interconnected instruments arranged to measure or control a process variable or both.

Manual Loading Station: A device having a manually adjustable output that is used to actuate one or more remote devices. Although the remote devices may be controller elements, the station does not provide switching between manual and automatic control modes of a control loop. The station may have internal gauges, lights, or other features, It is also known as a manual station, or a remote manual loader.

Measurement: The determination of the existence or magnitude of a variable. Measuring instruments include all devices used directly or indirectly for this purpose.

Pilot Light: A light that indicates which of a number of normal conditions of a system or device exists. It is unlike an alarm light, which indicates an abnormal condition. The pilot light is also known as *monitor* light.

Primary Element: That part of a loop or of an instrument that first senses the value of a process variable, and then assumes a corresponding predetermined and intelligible state or output. The primary element is separate from or integral with another functional element of a loop. The primary element is also known as a *detector* or a *sensor.*

Process: Any operation or sequence of operations involving a change of energy state, of composition, of dimension, or of any other property that may be defined with respect to a datum. The term *process* is used in this standard to apply to all variables other than instrument signals.

Process Variable: Any variable property of a process.

Relay: A device that receives information in the form of one or more instrument signals; modifies the information or its form, or both, if required; sends one or more resultant output signals; and is not designated as a controller, a switch or otherwise.

The term relay is also specifically applied to an electric switch that is remotely actuated by an electric signal. However for the purpose of standard, the term is not so restricted. The term is also applied to the functions performed by relays.

Scan: To sample each of a number of inputs intermittently. A scanning device may provide additional functions such as record or alarm.

Switch: A device that connects, disconnect, or transfers one or more circuits and is not designated as a controller, a relay, or a control valve. The term is also applied to the functions performed by switches.

Telemetry: The practice of transmitting and receiving the measurement of a variable for readout or other uses. The term is most commonly applied to electric signal systems.

Test Point: A process connection to which no instrument is permanently connected, but which is intended for temporary, intermittent, or future connection of an instrument.

Transducer: A general term for a device that receives information in the form of one or more physical quantities; modifies the information or its form or both, if required; and sends out a resultant output signal. Depending on the application, the transducer can be a primary element, a transmitter, a relay, a converter, or other device.

Transmitter: A device that senses a process variable through the medium of a primary element, and that has an output whose steady-state value varies only as a predetermined function of the process variable. The primary element may or may not be integral with the transmitter.

C.3 INSTRUMENT IDENTIFICATION

Each instrument can be identified by a tag number. It consists of a system of letters used to classify its functionality. To establish the identity of a loop for the instrument, a number is appended to the letters. This number is common to other instruments of the loop of which this instrument is a part. A suffix is sometimes added to complete the loop identification. Table C.1 shows the instrument identification scheme. A typical tag number for a temperature recording controller is also shown as an example in this table.

Table C.1 Instrument identification scheme

Tag number				
Functional identification			*Loop identification*	
First letter		*Succeeding letters*	*Loop number*	*Suffix*
Example:	T	R C	-2	A

The instrument tag number may include coded information such as a plant area designation, e.g. 6 in the tag number 6TRC-2A indicates the plant number.

Each instrument may be represented on the diagrams by a symbol. The symbol may be accompanied by an identification.

C.3.1 Functional Identification

The functional identification of the instrument consists of letters from Table C.2 and stands for the following:

First letter:	Measured or initiating variable
Succeeding letters:	Functions of the individual instrument

An exception to this rule is the use of the single letter L to denote a pilot light that is not part of an instrument loop.

Identification is made according to the function and not according to the construction of the instrument. For example, a differential pressure recorder used for the flow measurement is identified as FR.

In an instrument loop, the first letter of the identification shall be selected according to the measure or initiating variable and not according to the manipulated variable. Thus, a control valve varying flow according to the dictates of a level controller is an LV, not an FV.

The number of functional letters grouped for one instrument should be kept to a minimum (<4). All letters shall be uppercase.

C.3.2 Loop Identification

Each instrument loop has a unique number. An instrument common to two or more loops may have a separate loop number, e.g. FV-2A, FV-2B. An instrument that performs two or more functions may be designated by all its functions, e.g. a flow recorder FR-2, with a pressure pen PR-4 is designated as FR-2/PR-4.

A single sequence of loop numbers is used for all instrument loops of a project regardless of the first letter of the functional identification of the loops. A loop numbering sequence may begin with 1 or with any other convenient number, such as 301 or 1201, that may incorporate coded information such as plant area designation.

Instrument accessories, such as thermowell used with thermometer, if need tagging, are tagged individually and use the same loop number as that of the instrument they directly serve. Alternately, accessories may use the identical tag number as that of the instrument, but with the clarifying words added. For example, thermowell used with thermometer TI-9 may be tagged as TW-9 or TI-9 *thermowell.*

C.3.3 Function Designation for Relays

Table C.3 summarizes the function designation for relays.

C.3.4 Summary of Special Abbreviations

Table C.4 gives the summary of special abbreviations other than the instrument identification letters of Table C.2.

C.3.5 Drawings

The drawings as per ISA Standard S5.1 illustrate the symbols that are intended to depict instrumentation on flow diagrams and cover their applications to a variety of processes.

1. Instrument line drawings
2. General instrument symbols—baloons
3. Control valve body symbols
4. Actuator symbols
5. Symbols of self-actuated regulators, valves and other devices
6. Symbols for actuator action in the event of actuator power failure
7. Miscellaneous symbols
8. Primary element symbols
9. Function symbols
10. Miscellaneous systems

C.3.6 Method for Functional Diagrams

There are systems that are too complex to be shown with all their components on a flow diagram. Such a complex system may make use of an overall instrument designation. The operation of the system may be shown on a separate functional loop diagram that may also serve as the basis for subsequent development of a hardware diagram.

Figures C.1 (on pages 656 and 657) show two alternative methods of drawing flow diagrams of part of combustion control process in a power plant. Alternative 1 uses conventional tagging for all components of the system. Alternative 2 may be used if it is desired that the components mounted within the system cabinet not be tagged as usual.

C.3.7 P & I Diagrams

A Piping & Instrumentation (P & I) diagram shows a process using instrumentation symbols and identification standards explained above. Figure C.2 (on page 658) shows a P & I diagram of the control mechanism in a continuous flow stirred reactor (CSTR) in which both the reactor effluent and resultant by-product gases are relieved through the same outlet line. Reactor pressure is sensed and the overflow from the reactor is throttled to maintain the desired operating pressure. Process gas feed and process liquid feed streams are on the flow control.

Table C.2 Meanings of identification letters

	First letter		*Succeeding letters*		
Letter	*Measured or initiating variable*	*Modifier*	*Readout or passive function*	*Output function*	*Modifier*
A	Analysis		Alarm		
B	Burner flame		User's choice	User's choice	User's choice
C	Conductivity (Electrical)			Control	
D	Density (mass) or specific gravity	Differential			
E	Voltage (EMF)		Primary element		
F	Flow Rate	Ratio (Fraction)			
G	Gauging (Dimensional)		Glass		
H	Hand (manually initiated)				High
I	Current (Electrical)		Indicate		
J	Power	Scan			
K	Time or time schedule			Control station	
L	Level		Light (Pilot)		Low
M	Moisture or humidity				Middle or intermediate
N	User's choice		User's choice	User's choice	User's choice
O	User's choice		Orifice (Restriction)		
P	Pressure or vacuum		Point (test connection)		
Q	Quantity or event	Integrate or totalize			
R	Radioactivity		Record or print		
S	Speed or frequency	Safety		Switch	
T	Temperature			Transmit	
U	Multivariable		Multifunction	Multifunction	Multifunction
V	Viscosity			Valve, damper or louver	
W	Weight or force		Well		
X	Unclassified		Unclassified	Unclassified	Unclassified
Y	User's choice			Relay or compute	
Z	Position			Drive, actuate or unclassified final control element	

Table C.3 Function designations for relays

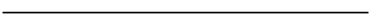
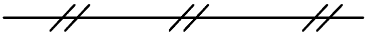
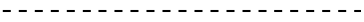
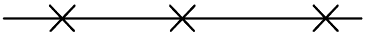
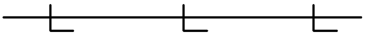
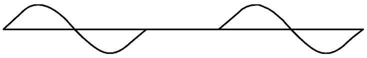
S.no.	*Symbol*	*Function*	
1.	1-0 or ON-OFF	Automatically connect, disconnect, or transfer one or more circuits provided that this is not the first such device in a loop.	
2.	Σ or ADD	Add or totalize (add and subtract)	
3.	Δ or DIFF	Subtract	
4.	± + −	Bias	
5.	AVG.	Average	
6.	% or 1:3 or 2:1 (typical)	Gain or attenuate (input : output)	
7.	×	Multiply	
8.	÷	Divide	
9.	$\sqrt{}$ or SQ. RT.	Extract a square root	
10.	x^n or $x^{1/n}$	Raise to power	
11.	$f(x)$	Characterize	
12.	1:1	Boost	
13.	> or HIGHEST (MEASURED VARIABLE)	High-select. Select the highest (higher) measured variable (not signal, unless so noted)	
14.	< or LOWEST (MEASURED VARIABLE)	Low-select. Select lowest (lower) measured variable (not signal, unless so noted)	
15.	REV.	Reverse	
16.	(a) E/P or P/I (typical)	For input sequences of the following:	
		Designation	**Signal**
		E	Voltage
		H	Hydraulic
		I	Current (electrical)
		O	Electromagnetic or Sonic
		P	Pneumatic
		R	Resistance (electrical)
	(b) A/D or D/A	For input sequences of the following:	
		A	Analog
		D	Digital
17.	$\int$	Integrate (time integral)	
18.	D or d/dt	Derivative or rate	
19.	1/D	Inverse derivative	
20.	As required	Unclassified	

Table C.4 Summary of special abbreviations (other than identification letters of Table C.2)

Abbreviation	*Meaning*
A	Analog signal
ADAPT.	Adaptive control mode
AS	Air Supply
AVG.	Average
C	Patchboard or matrix board connection
D	Derivative control mode Digital signal
DIFF.	Subtract
DIR.	Direct-acting
E	Voltage signal
ES	Electric Supply
FC	Fail Closed
FI	Fail Indeterminate
FL	Fail Locked
FO	Fail Open
GS	Gas Supply
H	Hydraulic signal
HS	Hydraulic Supply
I	Current (electrical) Interlock
M	Motor actuator
MAX.	Maximizing control mode
MIN.	Minimizing control mode
NS	Nitrogen Supply
O	Electromagnetic or sonic signal
OPT.	Optimizing control mode
P	Pneumatic signal Proportional control mode Purge or flushing device
R	Automatic-reset control mode Reset of failed-locked device Resistance (signal)
REV.	Reverse acting
RTD	Resistance (–type) Temperature Detector
S	Solenoid actuator
S.P.	Set Point
SQ. RT.	Square Root
SS	Stream Supply
T	Trap
WS	Water Supply
X	Multiply Unclassified actuator

C.3.8 Instrument Line Symbols

All lines shall be fine in relation to process piping lines.

- (i) Connection to process, or mechanical link, or instrument supply 
- (ii) Pneumatic signal or undefined signal for Process flow diagrams. 
- (iii) Electrical signal 
- (iv) Capillary tubing (filled system) 
- (v) Hydraulic signal 
- (vi) Electromagnetic or sonic signal (without wiring or tubing) 

Notes

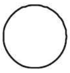
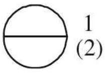
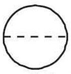
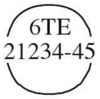
1. The power supply level also can be added to the instrument supply line, e.g. a 100 psig air supply line can be denoted as AS 100; a 24 V direct current supply can be denoted as ES 24 VDC.
2. The pneumatic supply symbol applies to a signal using any gas as the signal medium. If a gas other than air is used. The gas is identified by a note on the signal symbol or otherwise.
3. Air pressures for pneumatic controllers and transmission systems:

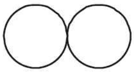
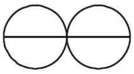
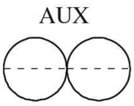
Operating pressure:	12 psi span—3 psi to 15 psi
ranges	24 psi span—3 psi to 27psi

Supply pressure: 12 psi span—not less than 18 psi and not greater than 20 psi
ranges 24 psi span—not less than 30 psi and not greater than 35 psi

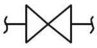
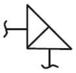
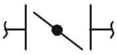
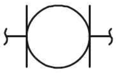
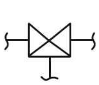
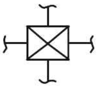
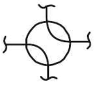
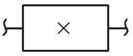
4. Electromagnetic phenomenon include heat, radio waves, nuclear radiation, and light.

C.3.9 General Instrument Symbols—Baloons

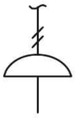
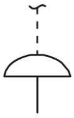
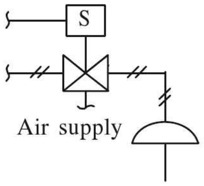
<table>
<tr>
<td>1.
Approximately 7/16″ diameter

Locally mounted</td>
<td>2.

Mounted on Board 1 or (Board 2). Board 2 may alternatively be designated by a double horizontal line instead of a single line, with the designation outside the baloon omitted.</td>
<td>3.

Mounted behind the board</td>
<td>4.

Locally mounted instrument with long Tag number. Alternatively, a circle can be enlarged.</td>
</tr>
<tr>
<td colspan="4">Instrument for single measured variable with any number of functions.</td>
</tr>
</table>

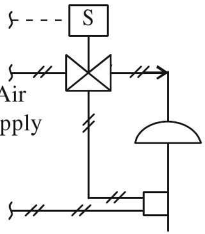
<table>
<tr>
<td>5.

Locally mounted</td>
<td>6.

Mounted on main board.</td>
<td>7.

Mounted behind the auxiliary board</td>
<td></td>
</tr>
<tr>
<td colspan="4">Instrument for two measured variables. Optionally, single-variable instrument with more than one function. Additional tangent baloons may be added as required.</td>
</tr>
</table>

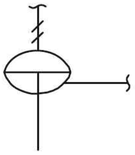
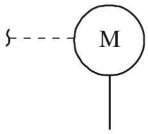
C.3.10 Control Valve Body Symbols

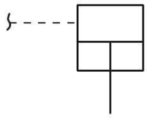
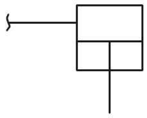
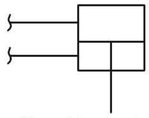
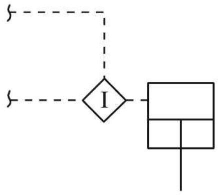
<table>
<tr>
<td>1.

Globe, gate or other in-line type not otherwise indentified</td>
<td>2.

Angle</td>
<td>3.

Butterfly, damper or louver</td>
<td>4.

Rotary plug or ball</td>
</tr>
<tr>
<td>5.

Three-way</td>
<td>6.

Four-way Alternative 1</td>
<td colspan="2">7.

Four-way Alternative 2</td>
</tr>
<tr>
<td colspan="4">8.

Unclassified. (It is expected that the type of body will be written in or adjacent to the symbol. Optionally, a number may be suffixed after X; e.g. X-1 to refer to item number 1 in a list of specific unclassified bodies.)</td>
</tr>
</table>

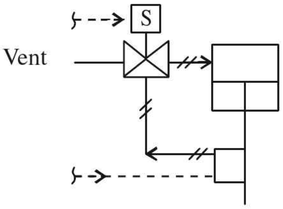
C.3.11 Actuator Symbols

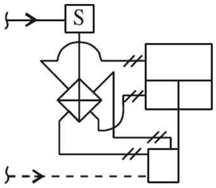
<table>
<tr>
<td>1.

Without positioner or other pilot</td>
<td>2.

Preferred for Diaphragm that is assembled with pilot</td>
<td>3.


Preferred alternative</td>
<td>4.


Optional alternative</td>
</tr>
<tr>
<td colspan="2">Diaphragm, spring opposed</td>
<td colspan="2">Diaphragm, spring opposed with positioner and overriding pilot valve that pressurizes the diaphragm when actuated</td>
</tr>
<tr>
<td>5.

Diaphragm, pressure balanced</td>
<td>6.


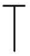
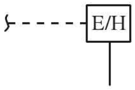
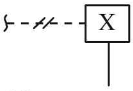
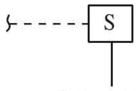
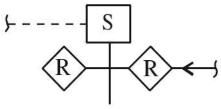
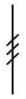
Rotary motor with electric signal</td>
<td colspan="2">7.

Preferred for any cylinder that is assembled with pilot so that the assembly is actuated by one controlled input</td>
</tr>
<tr>
<td colspan="2">8.

Single acting</td>
<td colspan="2">9.

Double acting</td>
</tr>
<tr>
<td colspan="4">Cylinder without positioner or other pilot</td>
</tr>
<tr>
<td>10.


Preferred alternative. A balloon with instrument tagging can be used instead of the interlock symbol ⟨I⟩</td>
<td>11.


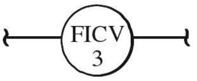
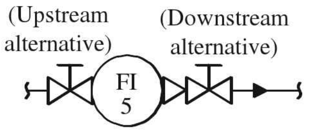
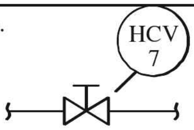
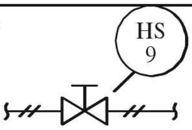
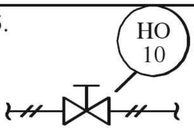
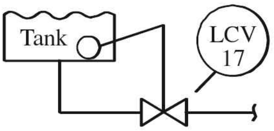
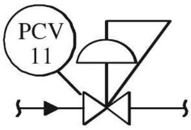
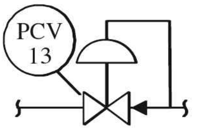
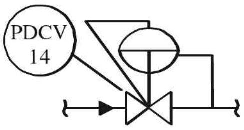
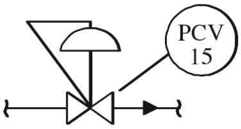
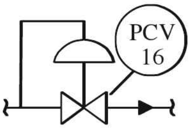
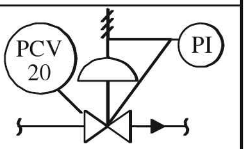
Single-acting cylinder (optional alternative)</td>
<td colspan="2">12.


Double-acting cylinder (optional alternative)</td>
</tr>
<tr>
<td colspan="4">Cylinder without positioner and overriding pilot valve</td>
</tr>
</table>

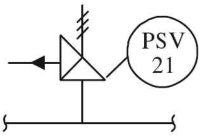
(*Contd.*)

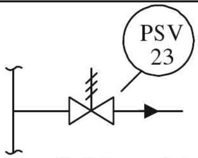
13.  Hand actuator (mounted on top side or the bottom of the actuated device as applicable)	14.  Electrohydraulic	15.  Unclassified (shown typically with pneumatic signal. It is expected that the type of actuator will be written in or adjacent to the symbol)
16.  Solenoid	17.  (Manual reset) (Remote reset) Latch-type actuator with reset (shown typically for solenoid actuator and typically with electric signal for remote reset with manual reset alternative)	18.  For pressure relief or safety valves only. Denotes a spring weight or integral pilot.

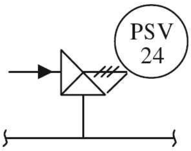
C.3.12 Symbols for Self-actuated Regulators, Valves and other Devices

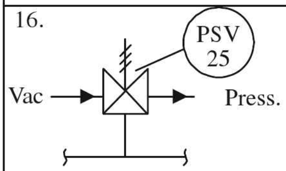
Flow	1.  Automatic regulator with integral flow indication	2.  (Upstream alternative) (Downstream alternative) FI 5 Indicating rotameter with integral manual throttle valve	
Hand	3.  Hand control valve in process	4.  Hand-actuated on-off switching valve in pneumatic signal line	5.  Manually adjustable restriction orifice in signal line
Level	6.  Level regulator with mechanical linkage		
Pressure	7.  Pressure reducing regulator, self-contained	8.  Pressure reducing regulator with external pressure tap	9.  Differential-pressure reducing regulator with external and internal pressure taps
	10.  Backpressure regulator, self-contained	11.  Backpressure regulator with external pressure tap	12.  Differential-pressure reducing regulator with external and internal pressure taps

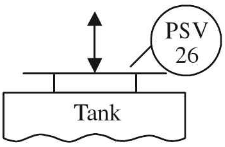
(*Contd.*)

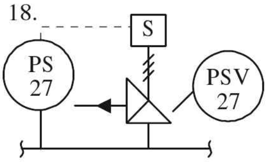
<table>
<tr><td rowspan="3">Pressure</td><td>13.


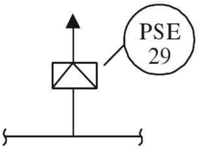
Pressure relief for safety valve, angle pattern, spring- or weight-loaded or with integral pilot</td><td>14.


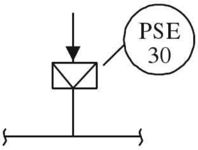
Pressure relief for safety valve, straight pattern, spring- or weight- loaded or with integral pilot</td><td>15.


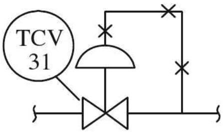
Vacuum relief valve, angle pattern, spring- or weight-loaded or with integral pilot</td></tr>
<tr><td>16.


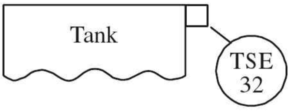
Pressure and vacuum relief valve, spring- or weight-loaded or with integral pilot</td><td>17.


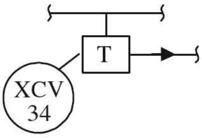
Pressure and vacuum relief manhole cover</td><td>18.


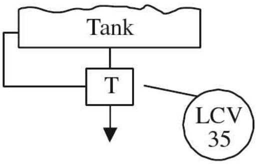
Pressure relief for safety valve, angle pattern, tripped by integral solenoid</td></tr>
<tr><td>19.


Rupture disk or safety head for pressure relief</td><td colspan="2">20.


Rupture disk or safety head for vacuum relief</td></tr>
<tr><td>Temperature</td><td>21.


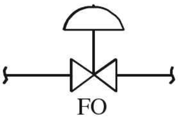
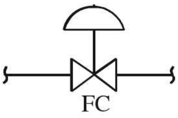
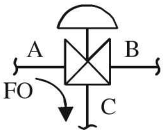
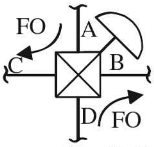
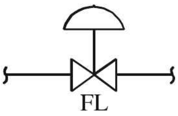
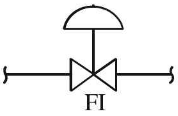
Temperature regulator, filled system type</td><td colspan="2">22.


Fusible plug or disk</td></tr>
<tr><td>Traps</td><td>23.


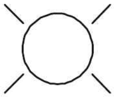
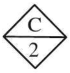
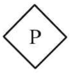
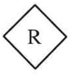
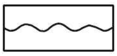
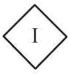
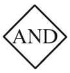
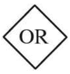
All traps other than ball-float type continuous drainers</td><td colspan="2">24.


Continuous drainer, ball-float type, with equalizing connection</td></tr>
</table>

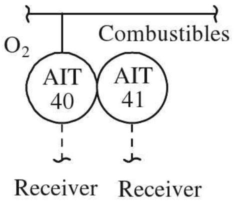
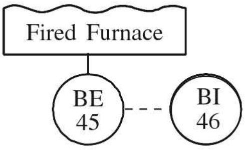
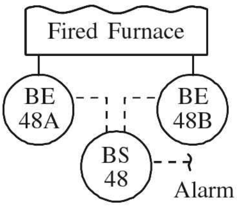
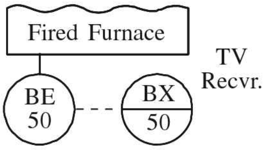
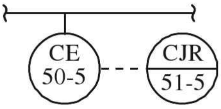
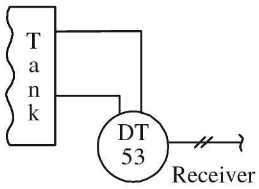
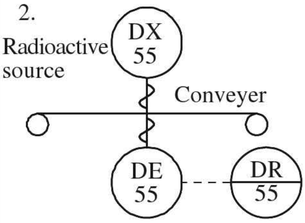
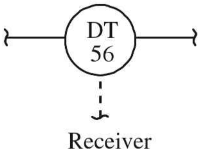
C.3.13 Symbols for Actuator Action in Event of Actuator Power Failure (Shown Typically for Diaphragm-actuated Control Valve)

1.  Two-way valve, Fail Open	2.  Two-way valve, Fail Closed	3.  Three-way valve, Fail Open to Path A-C
4.  Four-way valve, Fail open to paths A-C and D-B	5.  Any valve, Fail locked (Position does not change)	6.  Any valve, Fail indeterminate

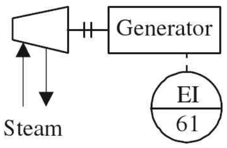
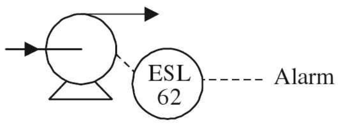
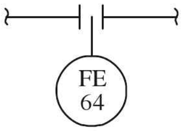
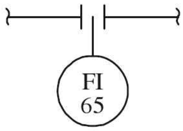
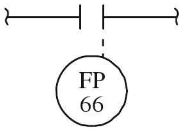
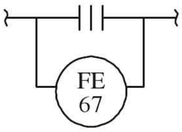
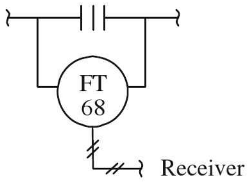
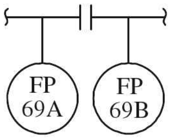
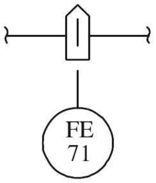
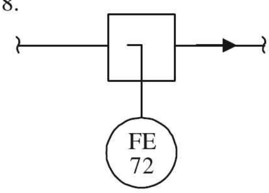
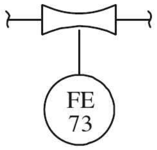
C.3.14 Miscellaneous Symbols

1. Approximately 7/16″ diameter  Pilot light	2. Approximately 7/16″ square  Board-mounted patch board or matrix board connection number 2	3. Approximately 1/4″ square  Purge or flushing device (means of regulating purge may be shown in place of symbol
4.  Rest for latch-type actuator	5.  Chemical seal	6.  Generalized—for undefined and or complex interlock logic
7.  Interlock is effective only if all inputs exist	8.  Interlock is effective if any one or more inputs exist	
Interlock		

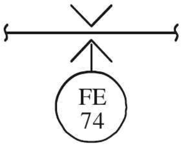
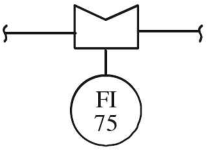
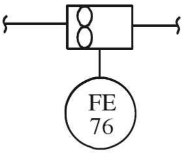
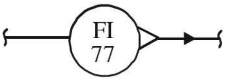
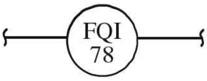
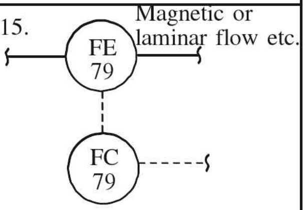
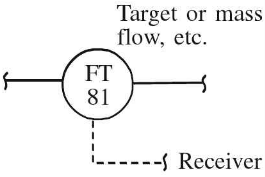
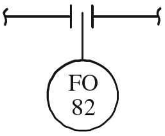
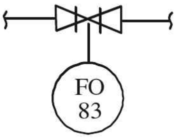
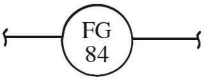
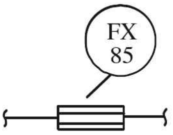
C.3.15 Primary Element Symbols

A Analysis	1.  O2 Combustibles AIT 40 AIT 41 Receiver Receiver Dual analysis indicating transmitter for oxygen and combustibles concentrations.		
B Burner flame	1.  One burner flame detector connected to analog-type flame intensity indicator	2.  Two burner flame sensors connected to common switch	1.  Television camera and receiver to view burner flame
C Conductivity	 Conductivity cell connected to point 5 of multipoint scanning conductivity recorder		
D Density or specific gravity	1.  Density transmitter, differential-pressure type, externally connected	2.  Radioactive type density element connected to density recorder on board	3.  Specific gravity transmitter, through flow type

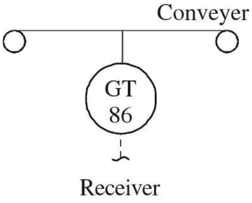
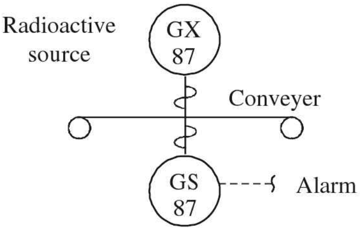
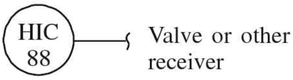
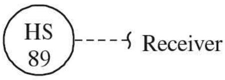
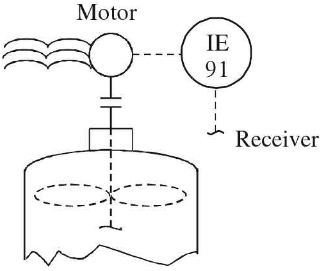
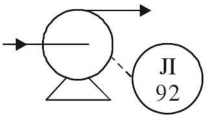
(*Contd.*)

E Voltage	1.  Indicating voltmeter connected to turbine-generator	2.  Low-voltage switch connected to pump motor	
F Flow rate	1.  Orifice plate with flange or corner taps	2.  Orifice plate with flange or corner taps connected to differential pressure type flow indicator	3.  Flange or corner tap test connections without orifice plate
	4.  Orifice plate with Vena Contracta, radius or pipe taps	5.  Orifice plate with Vena Contracta, radius or pipe taps connected to differential pressure type flow transmitter	6.  Vena Contracta, radius or pipe tap test connections without orifice plate
	7.  Orifice plate in quick-change fitting	8.  Pitot tube or Pitot-Venturi tube	9.  Venturi tube or Flow nozzle

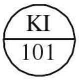
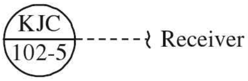
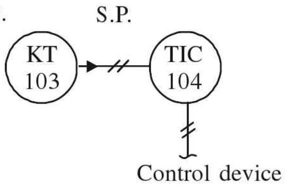
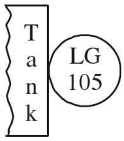
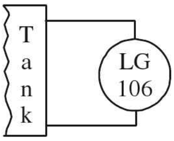
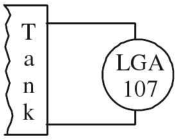
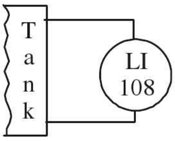
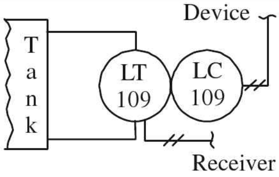
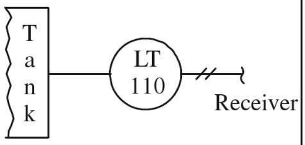
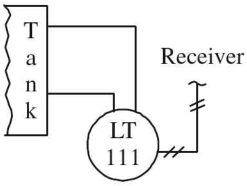
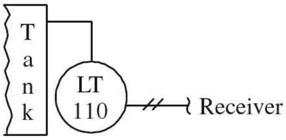
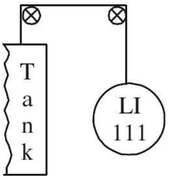
(*Contd.*)

F Flow rate	10.  Flume	11.  Weir	12.  Turbine- or propeller-type primary element
	13.  Rotameter-type flow indicator	14.  Positive displacement type flow totalizing indicator	15.  Unclassified primary flow element with external connection to flow controller (magnetic or laminar flow elements etc.)
	16.  Unclassified primary flow element integral with transmitter for mass flow or target type flow transmitters etc.)	17.  Fixed restriction orifice (orifice plate, capillary tube or multi-stage type, etc.) in process line	18.  Restriction orifice drilled in valve (instrument tag number may be omitted if valve is otherwise identified
	19.  Flow sight glass, plane or with paddle wheel, flapper, etc.	20.  Flow straightening vanes (use of tag number is optional. The loop number may be the same as that of the associated primary element)	

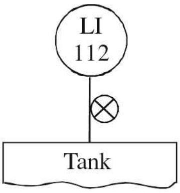
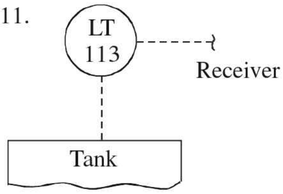
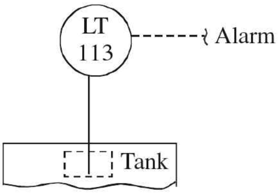
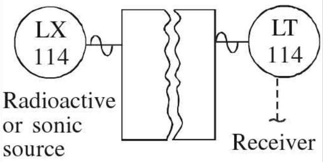
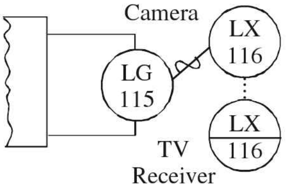
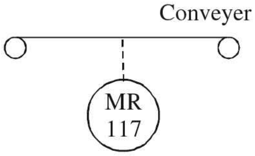
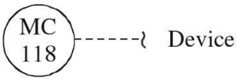
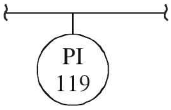
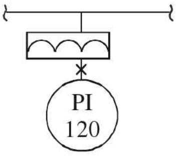
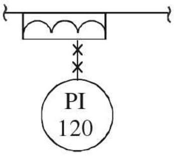
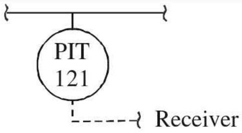
(*Contd.*)

G Voltage	1.  Roll-thickness transmitter	2.  Radioactive source GX 87 Conveyer GS 87 Alarm Plating thickness switch, radioactive type
H Hand	1.  Manual loading station with output gauge	2.  Hand-actuated electric switch
I Current	 Current transformer measuring current of electric motor	
J Power	 Indicating wattmeter connected to pump motor	

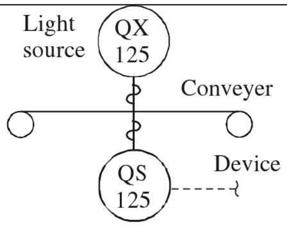
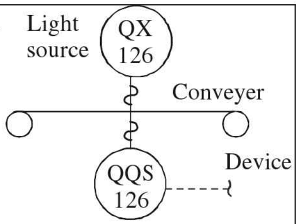
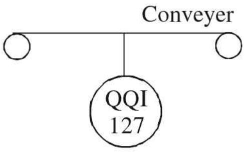
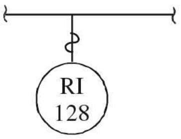
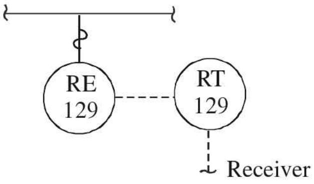
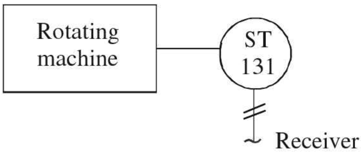
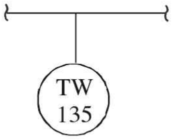
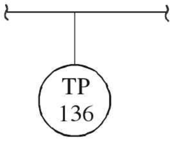
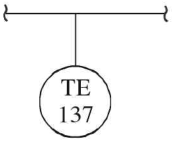
(*Contd.*)

K Time or time schedule	1.  Clock	2.  Multipoint on-off time-sequencing programmer point 5	3.  Time-schedule controller, analog type or self-contained function generator (for characterizing relay)
L Level	1.  Gauge glass integrally mounted on tank	2.  Gauge glass externally connected	3.  Water column with integral gauge glass and alarm whistle
	4.  Level indicator, magnetic-window type or float or displacer type	5.  Duplex level transmitter controller, external float or external displacer type	6.  Level transmitter, internal float type or displacer type, mounted on side of tank
	7.  Level transmitter, differential pressure type, externally connected with DIP tube	8.  Level transmitter, differential pressure type, mounted on tank	9.  Level indicator float type, with gauge board or tape indicator mounted below top of tank

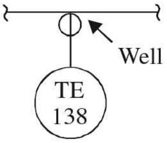
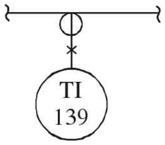
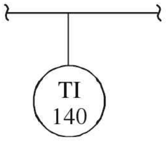
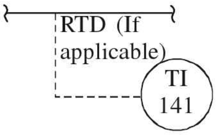
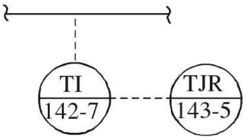
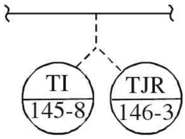
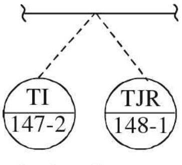
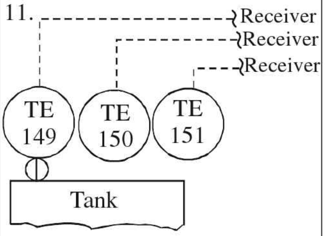
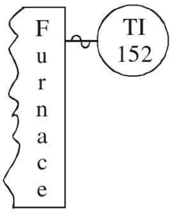
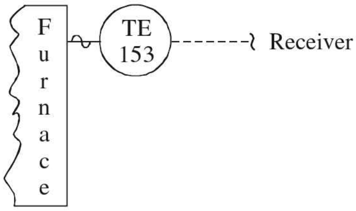
(*Contd.*)

L Level	10.  Level indicator, float type, with gauge board or tape indicator mounted above tank	11.  Capacitance or dielectric type level element connected to level transmitter	12.  Level switch, paddle-wheel or lever type, to measure level of solids
	13.  LX 114 LT 114 Radioactive or sonic source Receiver Radioactive or sonic type level transmitter with integral sensor	14.  Camera LX 116 LG 115 LX 116 TV Receiver Remote viewing of gauge glass by use of television camera and receiver	
M Moisture or humidity	1.  Moisture recorder (If there is a separate primary element, it shall be tagged ME-117)	2.  Self-contained humidity controller in room	
P Pressure or vacuum	1.  Pressure indicator, direct-connected	2.  Pressure indicator connected to chemical protector with filled system and with pressure lead line	3.  Pressure indicator connected to chemical protector with filled system linemounted
	4.  Pressure element, strain-gauge type, connected to pressure indicating transmitter (Tag strain-gauge PE-121).		

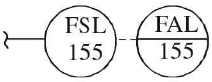
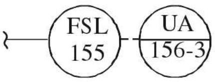
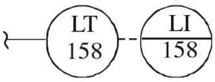
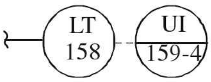
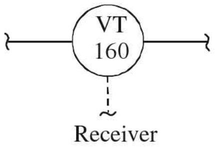
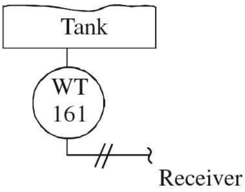
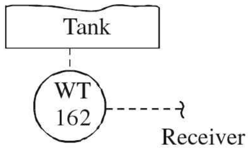
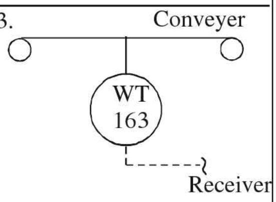
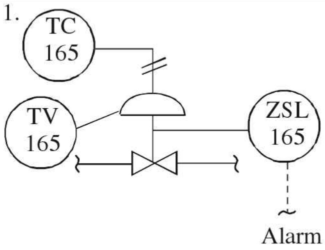
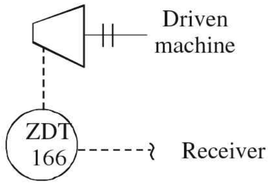
(*Contd.*)

Q Quantity or event (See section 9Q)	1.  Counting switch, photo-electric type, with switch action for each event	2.  Counting switch, photo-electric type, with switch action based on cumulative total	3.  Indicating counter, mechanical type
R Radioactivity	1.  Radioactivity indicator	2.  Radioactivity measuring element and transmitter	
S Speed or frequency	 Speed transmitter		
T Temperature	1.  Temperature test connection with well	2.  Temperature test connection without well	3.  Temperature element without well (element not connected to secondary instrument)

(*Contd.*)

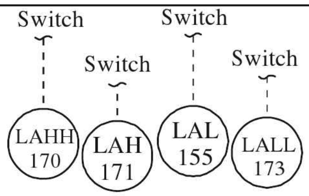
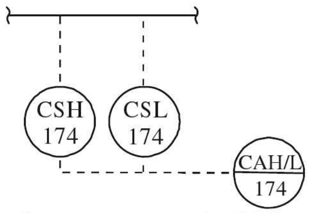
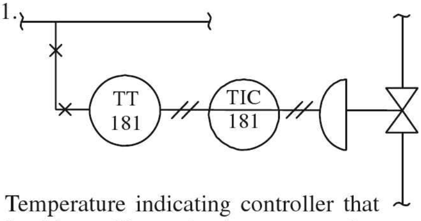
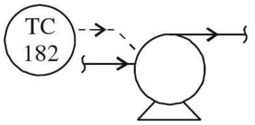
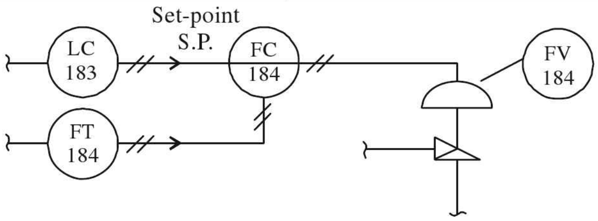
Temperature	4.  Temperature element with well (element not connected to secondary instrument)	5.  Filled-system-type temperature element with well	6.  Bimetallic type thermometer, glass thermometer or orther local unclassified temperature indicator
	7.  Thermocouple or RTD temperature element connected to temperature indicator (Tag element TE-141)	8.  Single thermocouple connected to multipoint temperature indicator having integral switch to permit recording on multipoint scanning temperature recorder (Tag temperature element TE-142-7)	9.  Single thermocouple parallel wired to multipoint temperature indicator and multipoint scanning temperature recorder (Tag temperature element TE-145-8/146-3)
	10.  Dual or duplex thermocouple connected to multipoint temperature indicator and multipoint scanning temperature recorder (Tag temperature element TE-147-2/148-1)	11.  Receiver Receiver Receiver TE 149 TE 150 TE 151 Tank Multi-element thermocouple for different elevations, with well in tank	12.  Thermal radiation type temperature indicator, self contained
	13.  Thermal radiation type temperature element		

(*Contd.*)

U Multivariable	1.  Alternative 1	2.  Alternative 2	3.  Alternative 1	4.  Alternative 2
	Low-flow switch actuating one point of multipoint multivariable alarm annunciator		Level signal received by one point of multipoint multivariable indicator	
V Viscosity	 Viscosity transmitter, through-flow type			
W Weight or force	1.  Weight transmitter, direct connected	2.  Strain gauge connected to separate weight transmitter, direct connected (Tag strain gauge WE-162)	3.  Weight-belt scale transmitter	
Z Position	1.  Limit switch that is actuated when valve closes to a predetermined position	2.  Turbine shell/rotor differential-expansion transmitter (Tag primary element ZDE-166)		

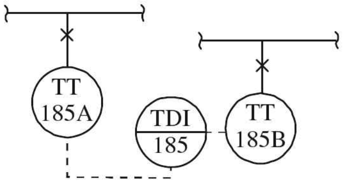
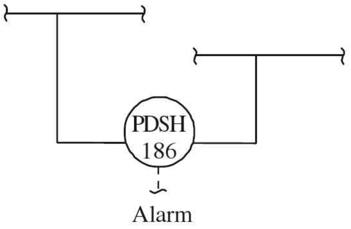
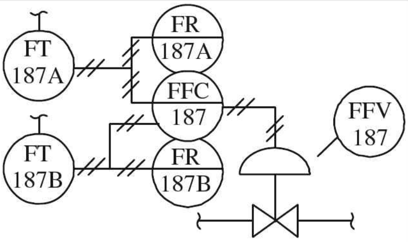
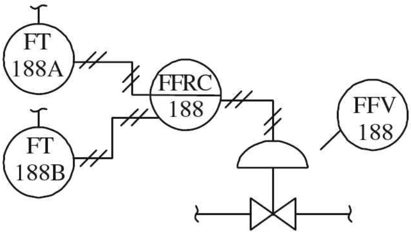
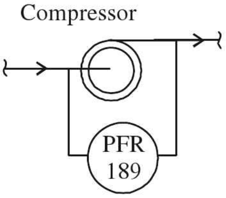
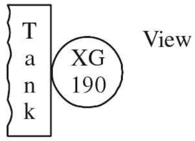
(*Contd.*)

C.3.16 Function Symbols

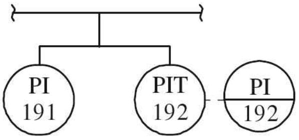
1. Alarm	1.  Switch Switch Switch Switch LAHH 170 LAH 171 LAL 155 LALL 173 Separate alarms for abnormal level—very high, high, low, and very low	2.  Common alarm annunciator for high and low conductivity
C* Control (see Section 5)	1.  Temperature indicating controller that is either self-contained or a control station/plug-in controller combination	2.  Self contained on-off room thermostat actuating circulating pump motor
	3.  Cascade control: Flow controller reset by level controller	

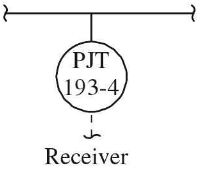
*It is expected that control modes will not be designated on a diagram. However, the following designations may be used outside the controller symbol, if desired, in combinations such as %, $\int$, D.

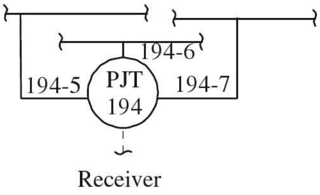
Control Mode	Designation	Control Mode	Designation
On-Off	1–0 or On-Off	Inverse derivative	1/D
Differential-gap, two position	Δ1–0 or On-Off	Optimizing	Opt. or Max. or Min. (as applicable)
Proportional	% or P		
Automatic reset, floating or integral	$\int$ or I	Adaptive	Adapt.
		Unclassified	As required
Derivative or rate	D or d/dt	Direct acting	Dir.
		Reverse acting	Rev.

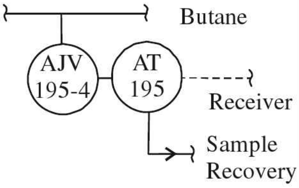
D Differential	1.  Differential-temperature indicator	2.  High-differential-pressure switch
E Primary element	See Section 8	
F Ratio	1.  FT 187A FR 187A FFC 187 FFV 187 FT 187B FR 187B Flow ratio controller with two pens to record flow	2.  Flow ratio controller with one pen to record flow-ratio
	3.  Direct-connected compression-ratio recorder	
G Glass (See Section 8)	 Sight glass for internal viewing	

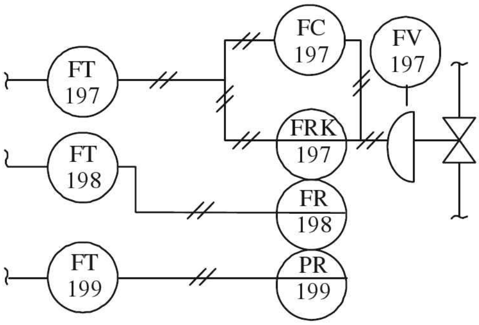
(*Contd.*)

<table>
<tr><td>H
High</td><td colspan="3">See Section 9 A, L</td></tr>
<tr><td>I
Indicate</td><td colspan="3">

Local pressure indicator and pressure indicating transmitter with common tap and board-mounted pressure receiver indicator</td></tr>
<tr><td>J
Scan
(See Sections 8T and 9R)</td><td>1.


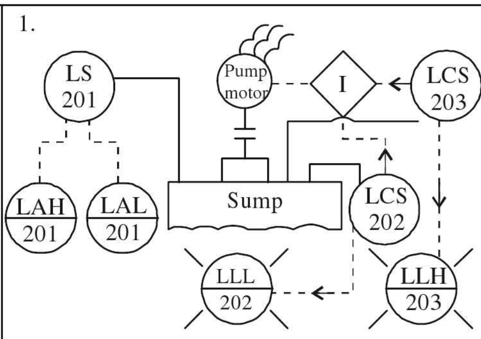
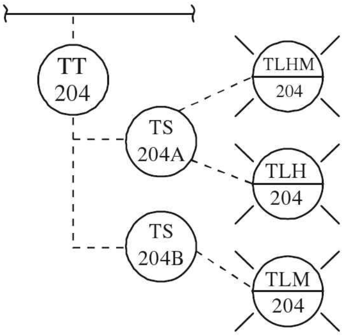
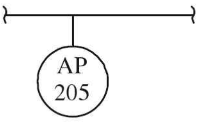
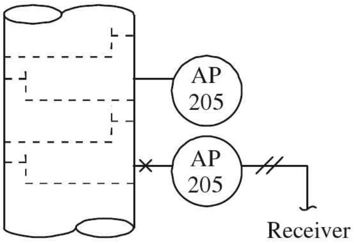
Pressure-scanning transmitter connected process point 4</td><td>2.


Pressure-scanning transmitter connected process points 5,6,7</td><td>3.


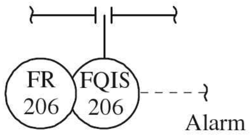
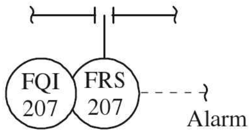
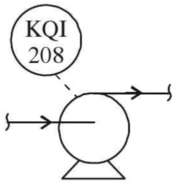
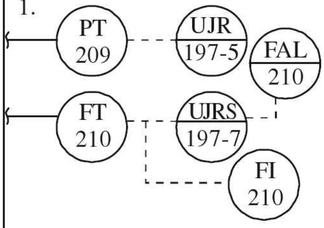
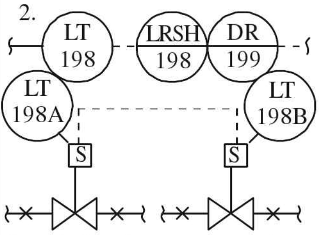
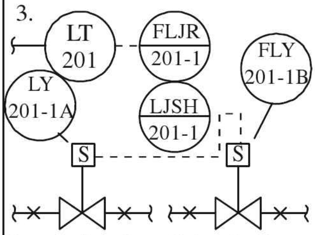
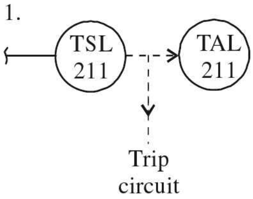
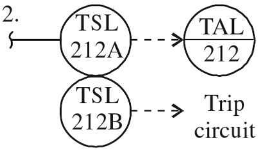
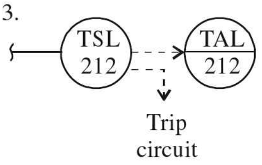
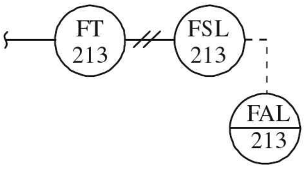
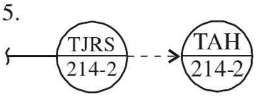
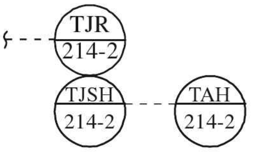
Sample line connected to butane-concentration transmitter through external sample scanning valve</td></tr>
<tr><td>K
Control station</td><td colspan="3">

Recording flow control station, board mounted, with additional flow and pressure pens and local controller</td></tr>
</table>

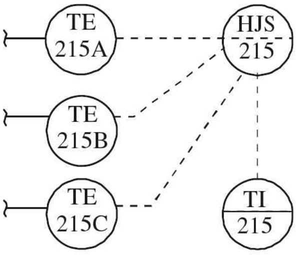
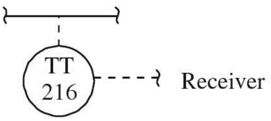
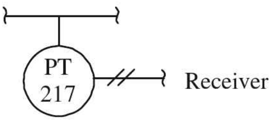
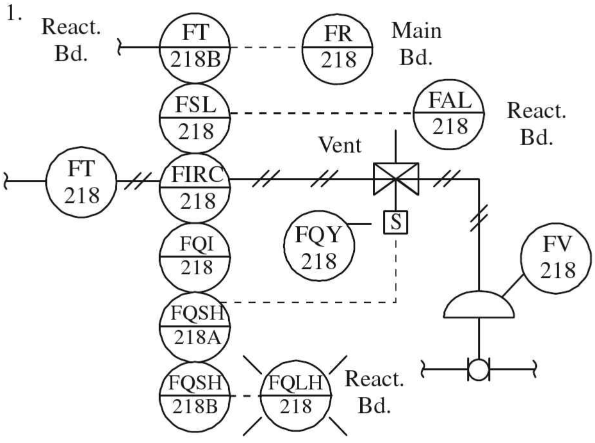
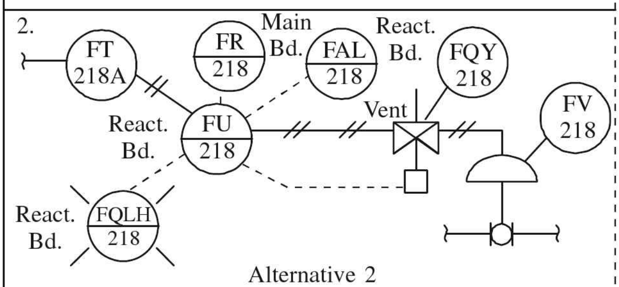
(Contd.)

L Light or low	1.  LS 201 Pump motor I LCS 203 LAH 201 LAL 201 Sump LCS 202 LLL 202 LLH 203 Differential-gap control of sump level through starting and stopping sump pump by LCS-202 and LCS-203 that also actuate High-and Low-level pilot lights. High- and Low-level alarms are actuated by LS-201	2.  TT 204 TLHM 204 TS 204A TLH 204 TS 204B TLM 204 Pilot lights to signal that temperature has risen as planned to give intermediate, High and Very High values
M Middle or Intermediate	See Section 9L	
O Orifice (Restriction)	See Section 5 and 8F	
P Point	1.  Analysis test sample point	2.  Distillation column with connection for alternative location of sensor of temperature transmitter

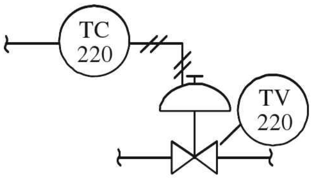
(*Contd.*)

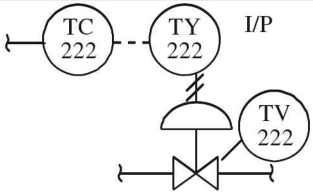
Q Integrate or totalize (See Section 8Q)	1.  Differential-pressure type flow meter with (1) recording of flow rate, (2) indication of integrated flow, and (3) switch actuated by integrated flow	2.  Differential-pressure type flow meter with (1) recording of flow rate, (2) indication of integrated flow, and (3) switch actuated by flow rate	3.  Running-time totalizer connected to pump motor
R Record or print	1.  PT 209 UJR 197-5 FAL 210 FT 210 UJRS 197-7 FI 210 Multipoint multi-variable data logger, UJR-197 with switch for point 7	2.  LT 198 LRSH 198 DR 199 LT 198A LT 198B S S Level signal to one pen of two-pen recorder, both pens record continuously. Level pen has switch that actuates two solenoid valves on High level	3.  LT 201 FLJR 201-1 FLY 201-1B LY 201-1A LJSH 201-1 S S Level signal to Point 1 of two-point level recorder. Points record alternatively. Point 1 has switch that actuates two solenoid valves on High level
S Switch for safety (See Section 5, Pressure and Temperature)	1. 	2.  Alternative 1	3.  Alternative 2
	Low temperature switch with one output for alarm and trip	Low temperature switch with two outputs for alarm and trip	
	4. 	5.  Alternative 1	6.  Alternative 2
	Receiver type flow switch actuating low-flow alarm	Multipoint temperature recorder with integral high-temperature switch for Point 2	

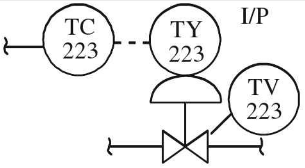
(*Contd.*)

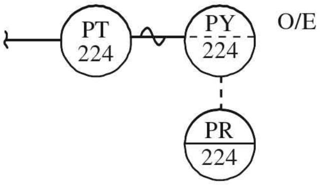
S (Contd.)	7.  One-point temperature indicator with separate manual selector switch	
T Transient	1.  Transmitter with thermoelectric input and electric output	2.  Transmitter with pneumatic output
U Multi-function	1.  React. Bd. FT 218B FR 218 Main Bd. FSL 218 FAL 218 React. Bd. Vent FT 218 FIRC 218 S FQI 218 FQY 218 FV 218 FQSH 218A FQSH 218B FQLH 218 React. Bd. Alternative 1 2.  FT 218A FR 218 Main Bd. FAL 218 React. Bd. FQY 218 Vent FV 218 React. Bd. FU 218 React. Bd. FQLH 218 Alternative 2	Flow measurement received by Reactor Board instrument that indicates and records flow, has switch to actuate low-flow alarm on Reactor Board, has a controller to actuate a control valve, retransmits the measurement to a recorder on Main Board, integrates the flow, indicates the integrated flow, has a switch to actuate a solenoid valve after a given integrated flow has passed and has another switch to actuate a pilot light after a greater integrated flow has passed.

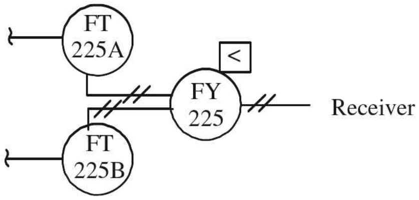
(*Contd.*)

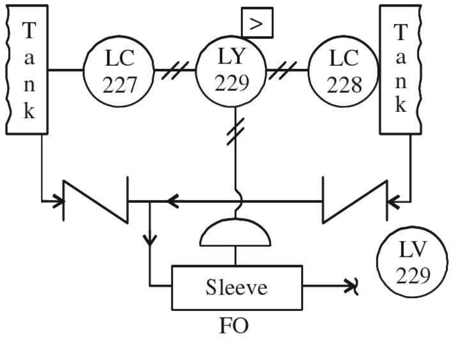
<table>
<tr><td>V
Valve, damper or louvre</td><td colspan="2">

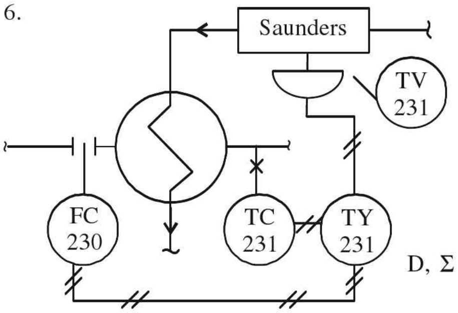
Control valve with mounted hand jack</td></tr>
<tr><td>W
Well</td><td colspan="2">See Section 9T</td></tr>
<tr><td rowspan="3">Y
Relay or compute</td><td>1. 

Electric current signal actuating pneumatic control valve with separately mounted electro-pneumatic converter</td><td>2. 

Electric current signal to pneumatic control valve furnished with electro-pneumatic converter, preferably, the converter symbol will be omitted (See Section 4).</td></tr>
<tr><td>3. 

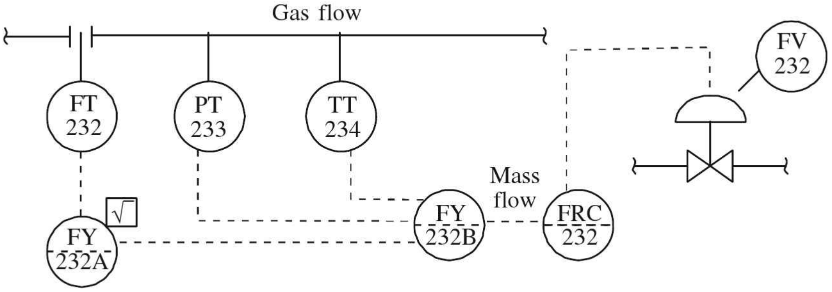
Pressure transmitter with radio (or laser) output, radio (or laser)-to-voltage converter and pressure recorder</td><td>4. 

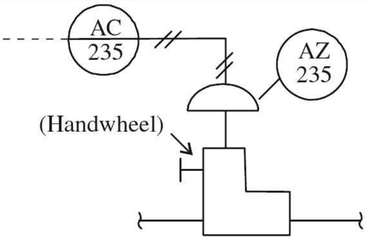
Selector relay whose output represents lower flow of FT-225A and FT-225B</td></tr>
<tr><td>5. 

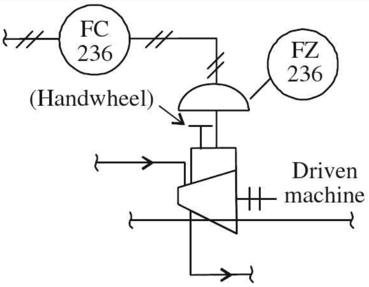
Tank outlet valve to open as required by the higher of the two levels</td><td>6. 

Temperature control with flow-anticipating relay</td></tr>
</table>

(*Contd.*)

<table>
<tr>
<td>Y
Relay or compute (Contd.)</td>
<td colspan="2">7.


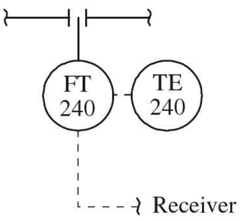
Mass-flow computation performed by Mass-flow computing relay FY-232B having linear inputs of uncompensated Flow, Pressure and Temperature</td>
</tr>
<tr>
<td>Z
Drive, actuate or unclassified final control element</td>
<td>1.


Proportioning pump with automatic and manual stroke control</td>
<td>2.


Turbine with automatic and manual speed changer</td>
</tr>
</table>

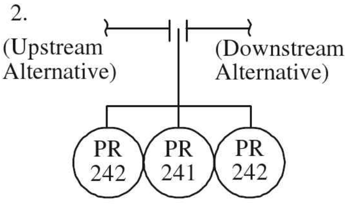
C.3.17 Miscellaneous Symbols

1.



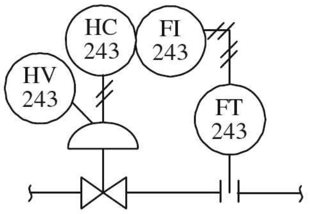
Flow transmitter with temperature element for temperature compensation

2.



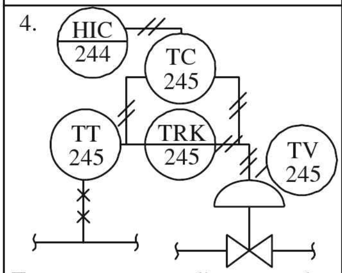
Flow recorder with pressure pen connected to tap differential pressure type flow recorder

3.



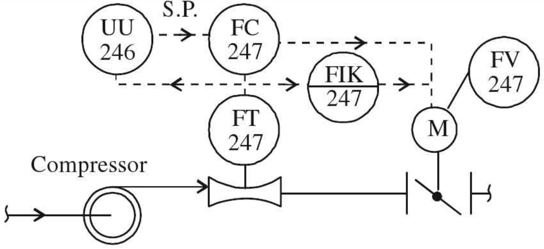
Manual loading station without output gauge and with flow receiver indicator

4.



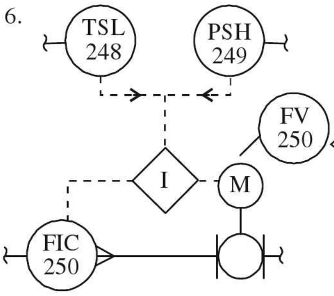
Temperature recording control station with local controller and remote manual gain adjustment

5.



Complex combustion control system, UU 246, of power boiler adjusting set point of flow control system for primary air to coal pulverizer

6.

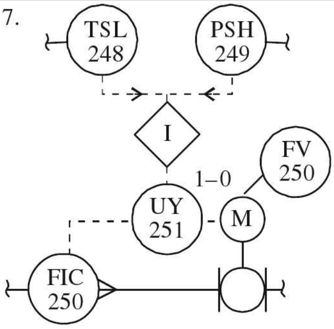


Preferred Method

Shall be used if interlock logic is undefined or complex. ◇OR may be substituted for ◇I if applicable.

◇I-1 may be substituted for ◇I to refer to interlock detail no. 1, to be drawn elsewhere. The outputs of TSL-248 and PSH-249 may enter the interlock box separately instead of jointly as shown.

7.



Optional Method

May be used if relay UY-251 exists and one desires to show it.

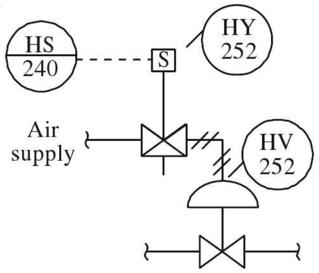
◇OR may be omitted, if desired.

And will be omitted if interlock logic is undefined or complex.

Flow control loop that may be blocked by interlock, actuated by either low temperature or high pressure

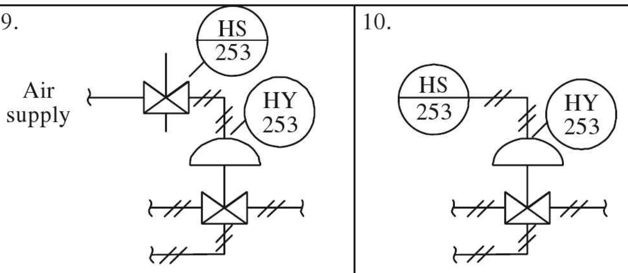
(*Contd.*)

8.



Solenoid pilot valve actuated by hand-actuated switch

9. 10.

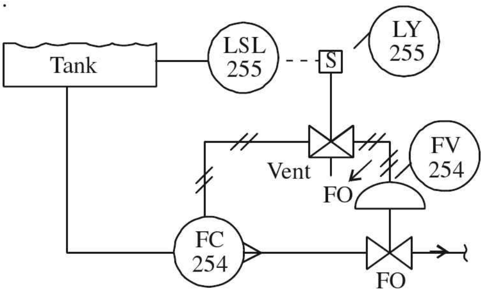


Alternative 1

Alternative 2

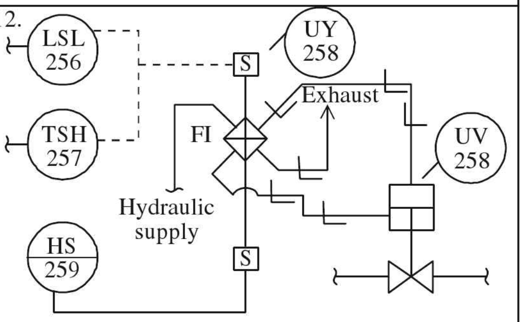
Pneumatic pilot valve actuated by remote hand-actuated switch

11.



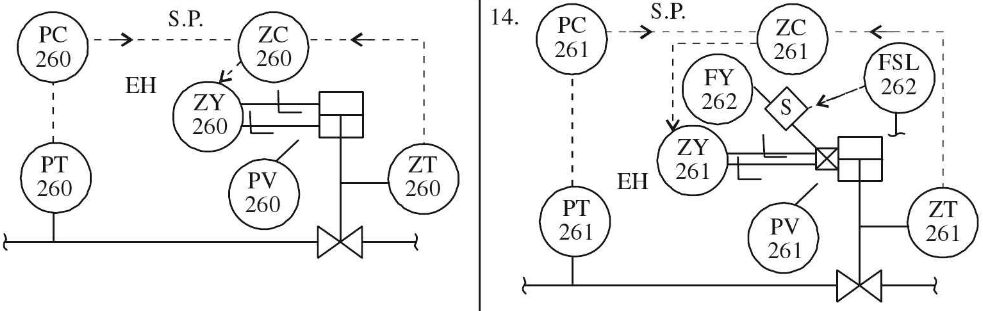
Flow control loop with low-level interlock actuated by level switch LSL-255

12.



Piston-actuated On-Off valve to be closed by low level or high temperature and to be reopened by pushbutton reset of double coil solenoid valve

13. 14.



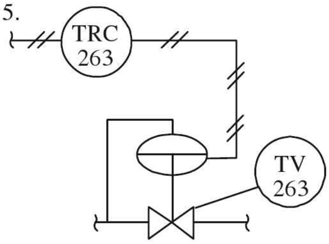
Without override

With low-flow through four-way solenoid-actuated pilot valve integral with control valve

Pressure controller acting through servo-valve acting as voltage-to-hydraulic converter to modulate double-acting-cylinder-actuated control valve having positioning system

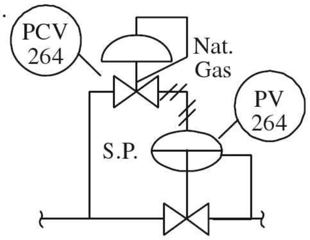
(*Contd.*)

15.



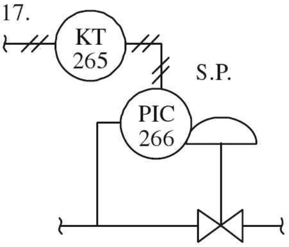
Pressure balanced valve

16.



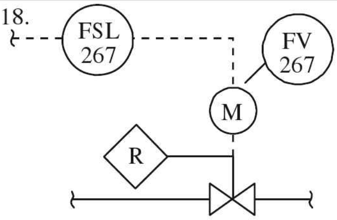
Self-actuated natural gas pressure regulator with remote set-point adjustment

17.



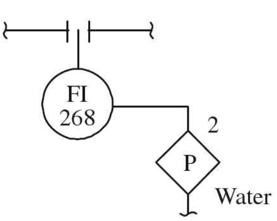
Time schedule controller varying set-point of valve-mounted pressure controller (Tag Valve assembly PIC/PV-2666)

18.



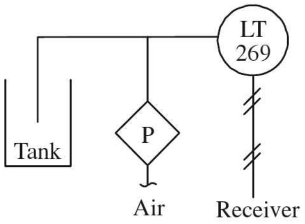
Valve that unlatches or latches when tripped and requires hand actuation of valve mounted reset mechanism

19.



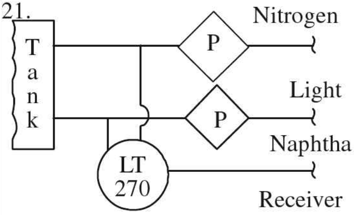
Differential-pressure type flow indicator with water purge for both leads (denoted by number 2)

20.



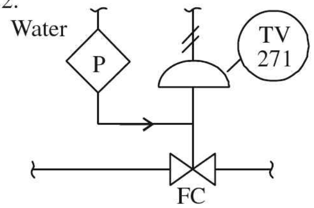
Level transmitter for open tank with DIP tube and air purge

21.



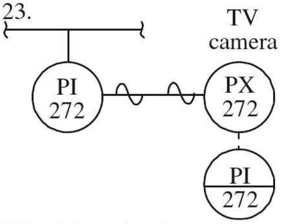
Differential-pressure type level transmitter with gas and liquid purges

22.



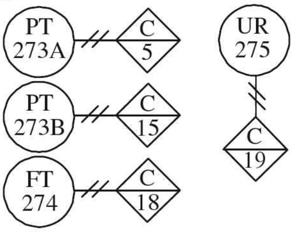
Control valve with water-sealed stuffing box. Valve closes on air failure

23.



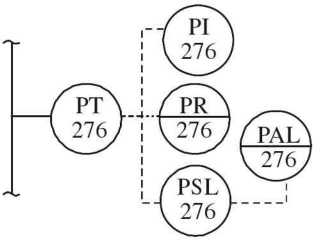
Television viewing of pressure gauge

24.



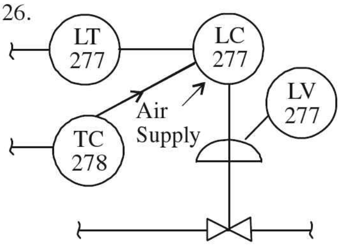
One-point recorder with patchboard input connection

25.



Electric transmitter connected to more than one receiver, regardless of type of electrical signal

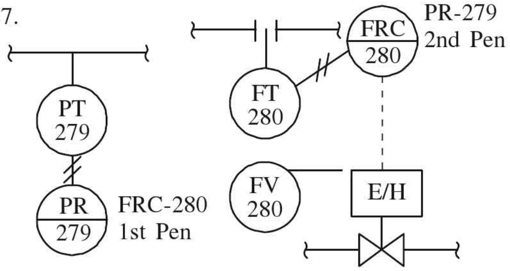
26.



Temperature controller output acting as air supply for level controller

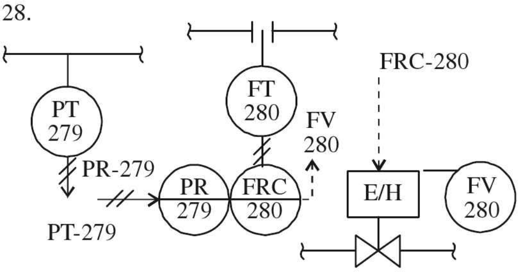
(*Contd.*)

27.



Alternative 1

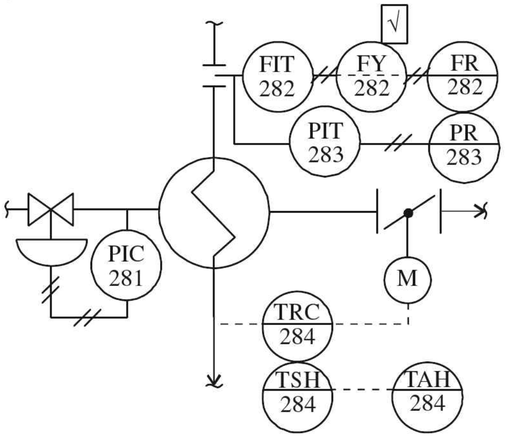
28.



Alternative 2

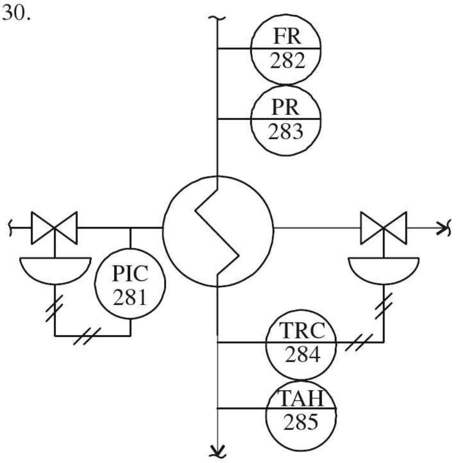
Interrelated instruments whose symbols are scattered on diagram. (A multi-point instrument, such as a datalogger, that is designated with point numbers on a diagram is not expected to have the symbols for the various points tied or referenced together)

29.



Typical symbolism for mechanical flow diagrams. Reflects actual hardware

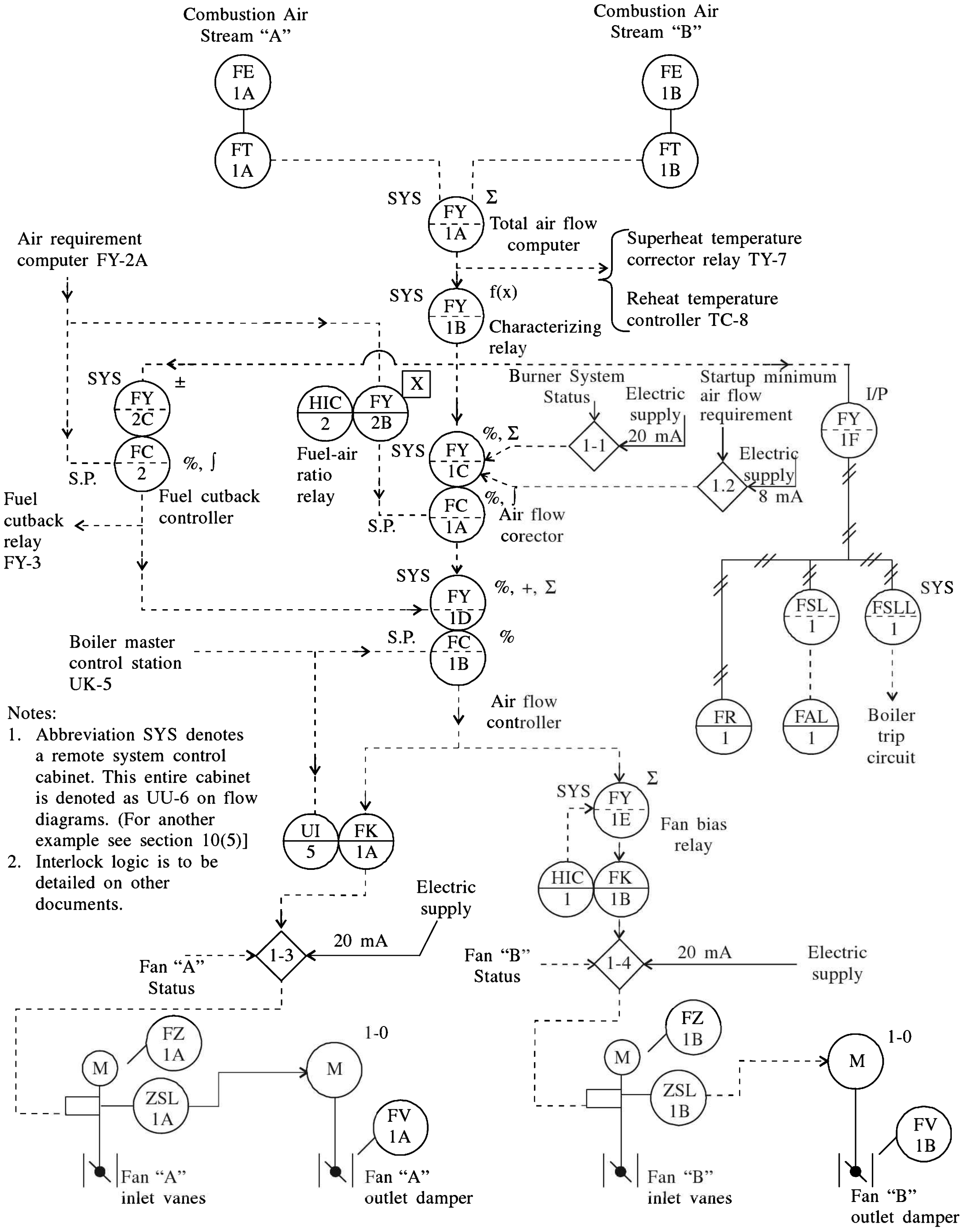
30.



Typical symbolism for process flow diagrams or other simplified flow diagrams that are supplemented by drawings showing instrument loop details.

Shows only the functions that are needed for operation of process. Proper kind of control valve may be arbitrarily chosen. Actual type of instrument signal may be shown, if desired. If option of simplified symbolism is chosen, it will be used consistently.

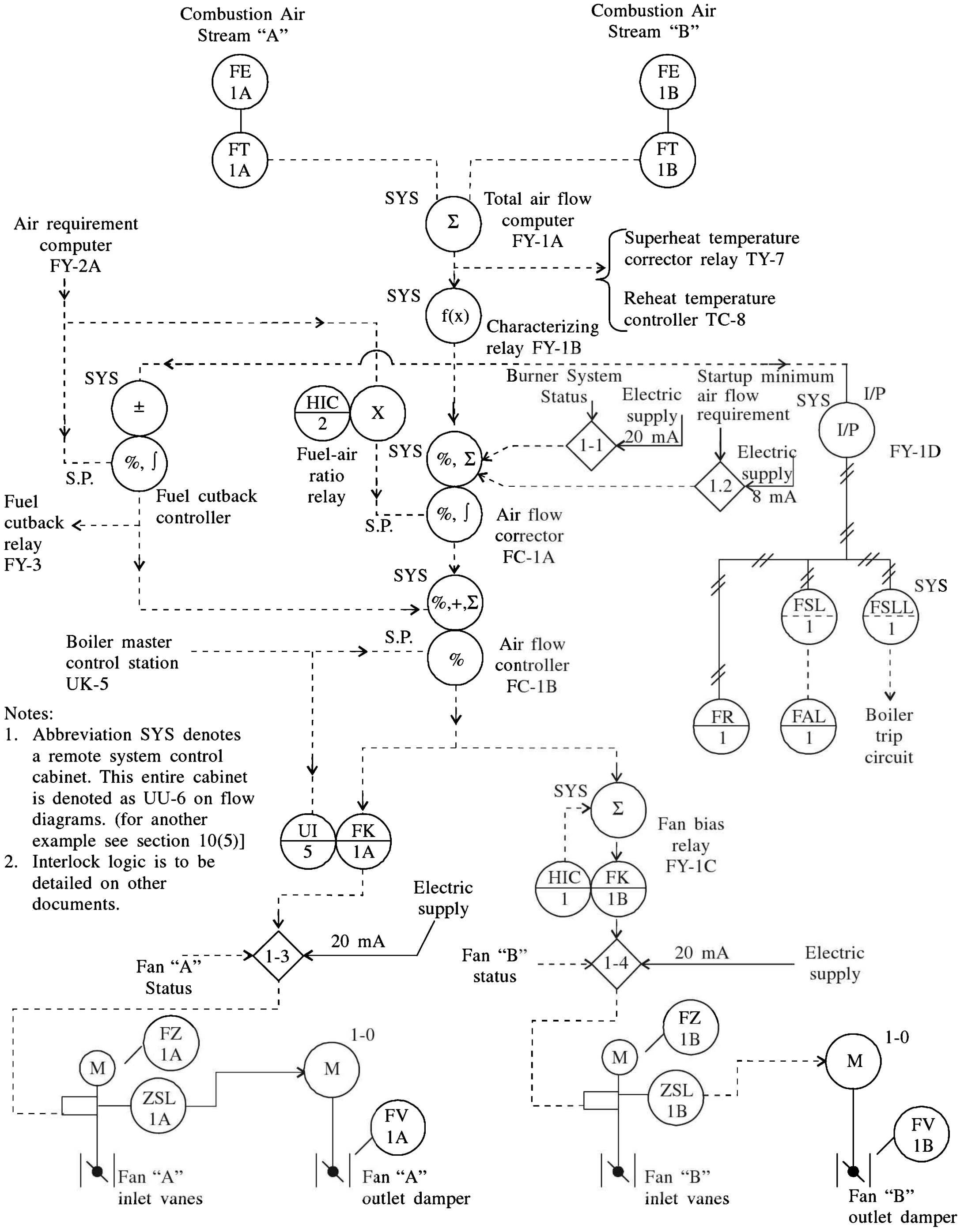
Methods of depicting instrument loops on flow diagrams



Notes:

1. Abbreviation SYS denotes a remote system control cabinet. This entire cabinet is denoted as UU-6 on flow diagrams. (For another example see section 10(5)]
2. Interlock logic is to be detailed on other documents.

Figure C.1(a) Alternative 1.



Notes:

1. Abbreviation SYS denotes a remote system control cabinet. This entire cabinet is denoted as UU-6 on flow diagrams. (for another example see section 10(5)]
2. Interlock logic is to be detailed on other documents.

Figure C.1(b) Alternative 2.

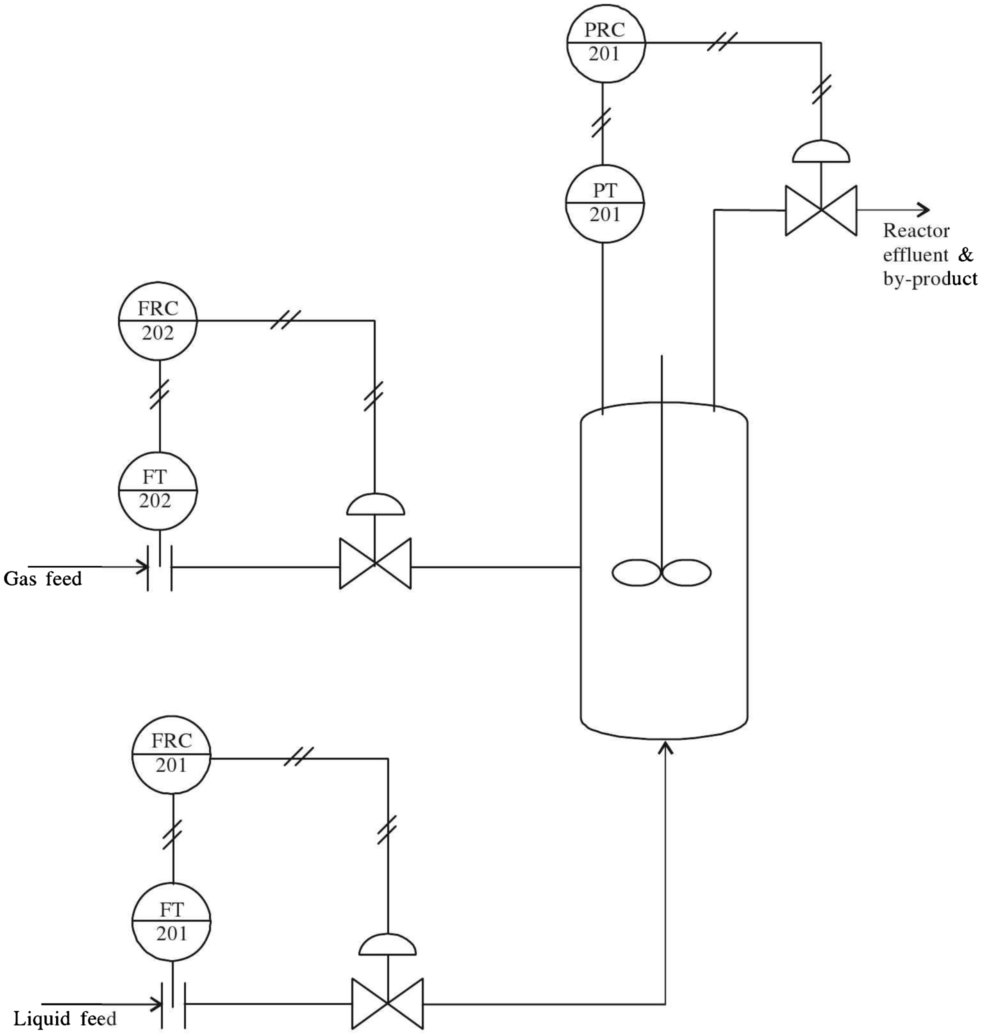


Figure C.2 P & I diagram of the control mechanism in a CSTR.

Appendix D

Typical Specifications of Instruments

MULTIMETER

TECHNICAL DATA

Note: Accuracy specifications apply to measurements made with the instrument in a horizontal position (meter facing upwards). Reference conditions: +25°C ±5°C; 45% to 75% relative humidity.

1. **DC Voltage**
   Ranges (full scale): 250 mV, 1.0 V, 2.5 V, 10 V, 25 V, 100 V, 250 V, 500 V and 1000 V
   Accuracy: ± 2% of full scale on all ranges
   Sensitivity: 20,000 Ω/V
2. **AC Voltage**
   Ranges (full scale): 2.5 V, 10 V, 25 V, 100 V, 250 V, 500 V and 1000 V
   Accuracy: ±3% of full scale on all ranges
   Sensitivity: 5000 Ω/V
   Frequency Response: See curve in Figure D.1.
3. **Ohms Conventional**
   Ranges: R × 1, R × 100, R × 1 kΩ and R × 10 kΩ
   Ohms Center: 6, 600, 6000 and 60 kΩ
   Max. Scale Reading: 1000 Ω (R × 1)
   Accuracy: ±2.5° of an arc on the R × 1 range; ±2.0° of arc on all other ranges. The nominal open-circuit voltage for all ranges up to R × 1 kΩ is 1.5 V. The R × 10 kΩ

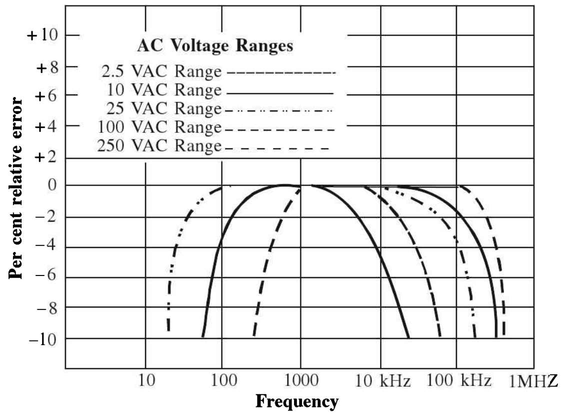


Figure D.1 Typical frequency response of various ac voltage ranges.

range has an open circuit voltage of 9 V. The maximum current drawn from the 1.5 V battery is 250 mA (R × 1 with test leads shorted).

4. Low Power Ohms

Ranges: R × 1 and R × 10
Ohms Center: 20 and 200 Ω
Max. Scale Reading: 1000 Ω (R × 1)
Accuracy: ± 2.5° of arc. The maximum open circuit voltage for the low power ohms ranges is 100 mV and the maximum measuring power is 0.125 mW. The battery quiescent current is 4.3 mA at R × 1 and 0.43 mA at R × 10.

5. DC Current

Ranges (full scale)	*Voltage drop*	*Accuracy*	*Internal resistance*
0–50 µA	250 mV	± 1.0% of FS	5000 Ω
0–0.5 mA	250 mV	± 2.0% of FS	500 Ω
0–5 mA	252 mV	± 2.0% of FS	50.4 Ω
0–50 mA	252 mV	± 2.0% of FS	5.04 Ω
0–500 mA	480 mV	± 2.0% of FS	0.8 Ω
0–5 A	250 mV	± 2.0% of FS	0.05 Ω

6. Output Jack

Voltage (AC)
Ranges: 2.5 V, 10 V, 25 V, 100 V, 250 V
Frequency Response: See curves in Figure D.1.

Decibels (dB):

Ranges (AC)	**Ranges (dB)**
2.5 V	Read Direct
10 V	Add 11 dB to reading
25 V	Add 19 dB to reading
100 V	Add 31 dB to reading
250 V	Add 39 dB to reading

ACCURACY: ±1.0 dB at the zero dB point

7. Rated Circuit-to-Ground Voltage*
(Float Potential): 1000 V AC/DC (1500 V peak) maximum

8. Readout: 4-1/2 inch, 50 μA (full scale) taut-band meter

*Per ANSI C39.5, April 1974 – "The specified voltage with respect to ground which may be safely and continuously applied to the circuits of an instrument."

9. Power Requirements: Two batteries: 1.5 V and 9 V cells.

10. Fuses: 1A, 250 V, Type 3 AG; 2A, Littlefuse Type BLS 600 V, Bussman BBS.

11. Dimensions: 5-1/4" wide × 7" long × 3-1/8" high (133 × 178 × 79 mm)

12. Weight: 1.14 kg

DIGIT DMM

DC VOLTAGE

Range	*Maximum reading*	*Best 6-1/2 digit accuracy*[1] ± (*% rdg + cnts*)		*Input resistance*
		% of reading	*Count error*	
30 mV	30.03000 mV	0.0040	365	10 GΩ
300 mV	303.0000 mV	0.0025	39	10 GΩ
3.0 V	3.030000 V	0.0017	6	10 GΩ
30.0 V	30.30000 V	0.0035	19	10 GΩ
300.0 V	303.0000 V	0.0050	6	10 GΩ

1. After 1 hr warm-up, integration time 100 PLC. Tcal is the temperature of the calibration environment between 18 and 28°C.

DC CURRENT

Range	*Maximum reading*	*Best 6-1/2 digit accuracy*[1] ± (*% rdg + cnts*)		*Input resistance*
		% of reading	*Count error*	
300 μA	303.0000 μA	0.02	104	1000 Ω
3 mA	3.030000 mA	0.02	104	100 Ω
30 mA	30.30000 mA	0.02	104	10 Ω
300 mA	303.0000 mA	0.07	204	1 Ω
1.0 A	1.000000 A	0.07	604	0.1 Ω

1. After 1 hr warm-up, integration time 100 PLC. Tcal is the temperature of the calibration environment between 18 and 28°C.

RESISTANCE (2 and 4-wire ohms)[2]

Range	*Maximum reading*	*Best* 6-1/2 *digit accuracy*[1] ± (*% rdg* + *cnts*)		*Current output*
		% of reading	*Count error*	
30 Ω	30.30000 Ω	0.0065	315	1 mA
300 Ω	303.0000 Ω	0.0045	34	1 mA
3 kΩ	3.030000 kΩ	0.0035	6	1 mA
30 kΩ	30.30000 kΩ	0.0035	6	100 μA
300 kΩ	303.0000 kΩ	0.0040	7	10 μA
3 MΩ	.030000 MΩ	0.0055	12	1 μA
30 MΩ	30.30000 MΩ	0.0250	80	100 nA
300 MΩ[3]	303.0000 MΩ	1.6	1000	100 nA
3.0 GΩ[3]	3.030000 GΩ	16.0	1000	100 nA

1. After 1 hr warm up, integration time 100 PLC. Tcal is the temperature of the calibration environment between 18 and 28°C.
2. For two-wire Ω, add 200 mΩ to count error specifications.
3. For two-wire Ω only. Accuracy is specified following autocal (ACAL), under stable conditions (±1°C).

Maximum Reading Rates (DCV, DCI, and Resistance up to 30 kΩ)[2]

Power Line cycles[3]	*Maximum # of digits*	*Readings per second*-60 Hz (50 Hz)		
		Auto zero On	*Auto zero Off*	*NMR*
0.0005	3-1/2	300	1350	0
0.005	4-1/2	280	1250	0
0.1	5-1/2	140 (128)	360 (312)	0
1.0	6-1/2	26 (22)	53 (45)	60 dB
10	7-1/2[1]	2.5 (2.0)	4.8 (4.0)	80 dB
100	7-1/2[1]	0.25 (0.2)	0.5 (0.4)	90 dB

1. Using Math HIRES mode.
2. Reading rates are specified with zero delay, fixed range, display off, and front panel off. The output is to internal reading memory using single integer format and internal timer.
3. Integration Time in Power Line Cycles (PLC).

Common Mode Rejection (dB)	(1 kΩ unbalance in low lead) DC ECMR 140 dB; AC ECMR: < 1 PLC, 76 dB; AC ECMR > 1 PLC 156 dB, for 50, 60 Hz ± .08%.
TRUE RMS ACV and (AC+DC)V	
Bandwidth	20 Hz to 100 kHz
Crest Factor	3.5 to 1 at full scale
Common Mode Rejection	(1 kΩ unbalance in low LO): > 76 dB, DC to 60 Hz
Accuracy (90 day)	Accuracy specified for sine wave inputs, > 10% of range. DC component < 10% of AC component after 1 hr warm-up and within one week of autocal. Integration time = 10 PLC. AC Band set to < 400 Hz. DC coupled mode requires 2 hour warm-up.

Range	*Maximum reading*	(100 Hz to 20 kHz) *Best* 5-1/2 *digit accuracy* ± (*% rdg + cnts*)				*Input impedance*
		AC coupled		*DC coupled*		
		% of reading	*Count error*	*% of reading*	*Count error*	
30 mV	32.50000 mV	0.13	116	0.17	364	1 MΩ ± 1% shunted by < 90 pF
300 mV	325.0000 mV	0.13	116	0.17	364	
3.0 V	3.250000 V	0.13	116	0.17	364	
30 V	32.50000 V	0.13	116	0.17	364	
300 V	303.0000 V	0.19	116	0.23	364	

TRUE RMS ACI and (AC + DC)I

Bandwidth 20 Hz to 100 kHz

Crest Factor 3.5 to full scale

Accuracy (90 day) Accuracy specified for sine wave inputs, > 10% of range. DC component < 10% of AC component after 1 hr warm-up and within one week of autocal. Integration time = 10 PLC. AC Band set to < 4000 Hz. DC coupled mode requires 2 hr warm-up.

Range	*Maximum reading*	(100 Hz *to* 20 kHz) *Best* 5-1/2 *digit accuracy* ± (*% rdg + cnts*)			
		AC coupled		*DC coupled*	
		% of *reading*	*Count error*	*% of reading*	*Count error*
30 mA	32.50000 mA	0.25	290	0.3	1600
300 mA	325.0000 mA	0.25	290	0.3	1600
1.0 A	1.000000 A	0.35	290	0.4	1600

Reading Rates (ACV and ACI)[1]

Power line cycles	*Maximum # of digits*	*Readings per Second* .60 Hz (50 Hz)	
		Input < 400 Hz (*Slow response*)	*Input* > 400 Hz (*Fast response*)
.0005	3-1/2	1	9.5
.005	4-1/2	1	9.5
.1	5-1/2	1 (1)	9.25 (9.2)
1	6-1/2	1 (1)	7.25 (6.9)
10	6-1/2	0.7 (0.65)	2.0 (1.7)
100	6-1/2	0.2 (0.17)	0.25 (0.2)

---

[1] Reading rates are specified with preprogrammed delays, fixed range, and autozero on.

FREQUENCY AND PERIOD

Measures the frequency or period of the ac component of the ac or dc coupled voltage or current input. The counter uses a reciprocal counting technique to give constant resolution independent of input frequency.

Input Impedance	Refer to AC voltage and current specifications.
Frequency Range	10 Hz to 1.5 MHz (voltage input) 10 Hz to 100 KHz (current input)
Period Range	.1 s to 667 ns (voltage input) .1 s to 3.33 μs (current input)
Sensitivity	10 mV or 100 μA (sinewave)
Triggering	Triggers and counts on zero crossings

Accuracy (1 year)

Frequency	*Period*	± % *of reading*
10–400 Hz	0.1–0.025 s	0.05
400 Hz–1.5 MHz	0.025 s–667 ns	0.01

Maximum Reading Rate	2.0 rdgs/s for integration time of 1 PLC, AC Band > 400 Hz, delay zero and math off, and fixed range.

MEMORY

2139 available bytes that can be partitioned into 3 segments, one devoted to storing measurements, one devoted to storing measurement subprograms, and one devoted to storing instrument states.

MATH FUNCTIONS

The 3457A performs the following math functions on the measurements—NULL, SCALE, OFFSET, RMS FILTER, SINGLE POLE FILTER, THERMISTOR LINEARIZATION, DB, DBM, % ERROR, PASS/FAIL LIMIT TESTING, and STATISTICS. Two math functions may be used at one time.

GENERAL SPECIFICATIONS

Operating Temperature	0 to 55°C
Warmup Time	one hour to all specifications except where noted
Humidity Range	95% R.H., 0 to 40°C
Storage Temperature	−40 to +75°C
Power	100/120/220/240 V ± 10%, 48 Hz to 66 Hz, 220 V, ±10%, 48 Hz to 66 Hz. Fused at 0.2 A (115 V) or 0.08 A (230 V). < 30 VA.
Size	89 mm H (without removable feet) × 425 mm W × 292 mm D (3.5″ × 16.75″ × 11.5″). Height (with removable feet): 100 mm (4″). Allow 76mm (3″) additional depth for wiring.
Net Weight	5.05 kg (11.1 lbs)
Shipping Weight	9.3 kg (20.5 lbs)

Plug-in Options	Eight two-wire armature relay channels and two current/actuator channels. Maximum voltage (terminal-to-terminal or terminal-to-chassis) 250 Vrms.
44491A Armature Relay Multiplexer Assembly Input Characteristics	Maximum current (per channel)—1.0 A DC or AC. Thermal Offset—3 μV. Closed channel resistance (end of relay life)—< 2 ohms. Maximum switching and measurement speed –33 channels/second.
44492A Reed Relay Multiplexer Assembly Input Characteristics	Ten two-wire reed relay channels Maximum voltage (terminal-to-terminal or terminal-to-chassis)—125 V peak. Thermal offset—3 μV. Closed channel resistance (end of relay life)—< 4 ohms. Specified for < 100 kHz ac volts and frequency operation. Maximum switching and measurement speed—300 channels/second.
44497A High Voltage Attenuator Assembly Input Characteristics	Two relay channels, channel 1 devoted to high voltage measurements. Maximum High-to-Low voltage of 1000 V DC or AC rms. Maximum Low-to-Earth voltage of 350 V Peak Non-destructive overload voltage of 1700 V Peak, 1200 V DC. Attenuator accuracy can be added to the 3457A range and function accuracy for total accuracy.
DC	0.030% of reading
20 Hz–1 kHz	2.8% of reading
1 kHz–10 kHz	10.0% of reading

Note: One year accuracy applies to Tcal ± 5%, NPLC = 1 or greater. Specifications are for Low-to-Earth voltage less than 0.1 times the High-to-Earth voltage.

DIGITAL STORAGE OSCILLOSCOPE

Parameter	**Value**
Total Channels	4
Sample Rate (all channels simultaneously)	5 GS/s
Real-time Bandwidth	500 MHz
Maximum Record Length per Channel	15,000 pts
Vertical Resolution	1 mV/div–10 V/div
Delta Time Measurement Accuracy	<50 ps @ 5 GS/s
Advanced Waveform DSP/Math	Std.
Histograms and Measurement Statistics	Std.
Standard Probes	4 P6243

Display Type	7 in. color display
Floppy Drive	Std.
GPIB, RS-232 & Centronics	Std.
VGA Output	Std. Colour
Hard Disk Drive	Opt.

TIME BASE SYSTEM

Time Bases—Main and delayed

Time/div Range—200 ps/div to 10 s/div

Time Base Accuracy—Over any interval >1 ms ±100 ppm

Time Interval Measurement Accuracy—$\pm[(0.2/\text{sample rate}) + (100\text{ ppm} \times |\text{Reading}|)]$ single shot. (approximately 50 ps @ 5 GS/s)

Record Length per Channel—500 to 15,000 pts

Trigger Jitter—8 ps RMS (typical)

Pre-Trigger Position—0% to 100% of record.

Channel to Channel Deskew Range—±25 ns.

VERTICAL SYSTEM

Vertical Resolution—8-bits (>11-bits with averaging).

Vertical Sensitivity—1 mV/div to 10 V/div.

Maximum Input Voltage—300 V CAT II; ±400 V peak. Derate at 20 dB/decade above 1 MHz.

DC Gain Accuracy—±1.5%.

Position Range—±5 divs.

Offset—±1 V from 1 to 99.5 mV/div, ±10 V from 100 mV to 995 mV/div, ±100 V from 1 V to 10 V/div.

Bandwidth Selections—20 MHz, 250 MHz, and Full.

Input Impedance Selections—1 megaohm in parallel with 10 pF, or 50 ohm (AC and DC coupling).

Input Coupling—AC, DC or GND.

AC Coupled Low Frequency Limit—<10 Hz when AC, 1 megaohm coupled. <200 kHz when AC, 50 ohm coupled.

Channel Isolation—>100:1 at 100 MHz and >30:1 at BW for any two channels having equal V/div settings.

ACQUISITION MODES

Peak Detect—High frequency and random glitch capture. Captures glitches of 1 ns using acquisition hardware at all real-time sampling rates.

Sample—Sample data only.

Envelope—Max/min values acquired over one or more acquisitions.

Average—Waveform data from 2 to 10,000 waveforms (selectable) is averaged.

Single Sequence—Use RUN/STOP button to capture a single triggered acquisition at a time, which may be automatically saved to NVRAM with AutoSave.

TRIGGERING SYSTEM

Trigger Types

EDGE (Main and Delayed)

Conventional level-driven trigger. Positive or negative slope on any channel or rear panel auxiliary input. Coupling selections: DC, AC, noise reject, HF reject, LF reject.

LOGIC (Main)

PATTERN: Specifies a logical combination (AND, OR, NAND, NOR) of the four input channels (high, low, don't care). Trigger when pattern stays true or false for a specified time.

STATE: Any logical pattern of channels 1, 2, and 3 plus a clock edge on channel 4. Triggerable on rising or falling clock edge.

SETUP/HOLD: Trigger on violations of both setup time and hold time between clock and data which are on two input channels.

PULSE (Main)

GLITCH: Trigger on or reject glitches of positive, negative, or either polarity. Minimum glitch width is 1.0 ns with 200 ps resolution.

RUNT: Trigger on a pulse that crosses one threshold but fails to cross a second threshold before crossing the first again.

WIDTH: Trigger on width of positive or negative pulse either within or out of selectable time limits (1 ns to 1 s).

SLEW RATE: Trigger on pulse edge rates that are either faster or slower than a set rate. Edges can be rising, falling, or either.

TIMEOUT: Trigger on an event which remains high, low, or either, for a specified time period, selectable from 1 ns to 1 s, with 200 ps resolution.

VIDEO (Optional)

Trigger on a particular line of individual, odd/even, or all fields. Trigger on a specific pixel of a line by using the video trigger with delay by events. Choose positive or negative horizontal sync polarity.

525/NTSC: Choose monochrome or colour (studio-quality NTSC) sync formats. 625/PAL: Choose colour or monochrome (studio-quality PAL) sync formats. HDTV: Choose from 1125/60, 1050/60, 1250/50, and 787.5/60 HDTV formats.

Trigger Bandwidth (Edge Type)—500 MHz

Main Trigger Modes—Auto, Normal, Single.

Delayed Trigger—Delay by time, events, or events and time.

Delay by Time Range—16 ns to 250 s.

Delay by Events Range—2 to 9,999,999 events.

External Trigger Input—Input Impedance: ≥1.5 kΩ; Max. input voltage: ±20 V (DC + peak AC).

DISPLAY CHARACTERISTICS

Waveform Style—Dots, vectors, variable persistence from 32 ms to 10 s, infinite persistence, and intensified samples.

Colour—Standard palettes and user-definable colour for waveforms, text, graticules, and cursors. Measurement text and cursor colours matched to waveform. Waveform collision areas highlighted with different colour. Statistical waveform distribution shown with colour grading through variable persistence.

Colour Grading—With variable persistence selected, historical timing information is represented by temperature or spectral colour scheme.

Graticules—Full, grid, cross-hair, frame, and NTSC and PAL (with video trigger option).

Format—YT and XY.

Resolution—640 horizontal by 480 vertical displayed pixels (VGA).

Colour CRT Monitor—7 in. diagonal NuColour$^{\text{TM}}$ liquid crystal full-colour shutter, 256 levels.

MEASUREMENT SYSTEM

Automatic Waveform Measurements

Period, frequency, + width, – width, rise time, fall time, + duty cycle, – duty cycle, delay, phase, burst width, high, low, max, min, peak to peak, amplitude, + overshoot, – overshoot, mean, cycle mean, RMS, cycle RMS, area, cycle area, extinction ratio (ratio, dB, %), and mean optical power. Continuous update of up to four measurements on any combination of waveforms.

Measurement Statistics—Display minimum and maximum or mean and standard deviation on any displayed single-waveform measurements.

Thresholds—Settable in percentage or voltage.

Gating—Any region of the waveform may be isolated for measurement using vertical bars.

Snapshot—Performs all measurements on any one waveform showing results from one instant in time.

Cursor Measurements—Absolute, Delta: Volts, Time, Frequency, and NTSC IRE and line number (with video trigger option).

Cursor Types—Horizontal bars (volts), vertical bars (time); operated independently or in tracking mode.

Histogram Measurements—Mean, median, standard deviation, hits, waveform count, peak hits, peak to peak, % mean ± 1, 2, and 3 standard deviations.

WAVEFORM PROCESSING

Waveform Functions—Interpolation (sin(x)/x or linear), Average, Envelope, Auto Setup.

Advanced Waveform Functions—FFT, Integration, Differentiation, Waveform (math or acquired) Limit Testing.

Arithmetic Operators—Add, Subtract, Multiply, Divide, Invert.

Autoset—Single-button, automatic setup on selected input signal for vertical, horizontal, and trigger systems.

Waveform Limit Testing—Compares incoming or math waveform to a reference waveform's upper and lower limits.

Waveform Histograms—Both vertical and horizontal histograms, with periodically updated measurements, allow statistical distributions to be analyzed over any region of the signal.

ZOOM CHARACTERISTICS

The zoom feature allows waveforms to be expanded or compressed in both vertical and horizontal axes. Allows precise comparison and study of fine waveform detail without affecting ongoing acquisitions. When used with Average acquisition modes, Zoom provides an effective vertical dynamic range of 1000 divisions or 100 screens.

Dual Window Zoom—Dual graticules simultaneously show selected and zoomed waveforms. Up to two zoom boxes show areas on the selected trace that are being magnified, and the two magnified areas can be overlapped for quick comparison. Colour of zoomed trace matches selected trace.

COMPUTER INTERFACE

GPIB (IEEE-488.2) Programmability—Full talk/listen modes. Control of all modes, settings, and measurements.

HARDCOPY

Printer—HP Thinkjet, Deskjet, Laserjet, Epson, Interleaf, PostScript, TIFF, PCX, BMP, DPU411/412, RLE.

Plotter—HPGL.

Data—MathCad, spreadsheet formats.

Hardcopy Interface—Centronics and RS-232 (talk only).

DATA STORAGE

Standard—Internal MS-DOS 1.44 Mbyte floppy, external Iomega Zip$^{\text{TM}}$ drive support.

Optional—Internal hard disk drive.

Non-volatile Storage for Setups—10 front panel setups.

Floppy Disk Drive—Store reference waveforms, setups, and image files on 3.5 in. 1.44 MB or 720 K DOS-format floppy disk.

Iomega Zip$^{\text{TM}}$ Drive Compatible—Compatible for waveform, hardcopy, and front panel setup file transfer to Iomega Zip$^{\text{TM}}$, ZipPlus$^{\text{TM}}$, and Zip250$^{\text{TM}}$ Drive.

POWER REQUIREMENTS

Line Voltage Range—90 to 250 V RMS.

Line Frequency—45 to 440 Hz.

Power Consumption—300 W max.

ENVIRONMENTAL AND SAFETY

Temperature—Operating: +4°C to +45°C (floppy not used), +10°C to +45°C (floppy in use). Non-operating: –22°C to +60°C.

Humidity—Operating: 20% to 80% RH at or below +32°C.

Altitude—Operating: 15,000 ft. (hard disk not used), 10,000 ft. (hard disk in use). Non-operating: 40,000 ft.

Electromagnetic Compatibility—UL1244.

Safety—UL 3111-1, CSA1010.1, EN61010-1, IEC61010-1.

LCR METER

Introduction

The QuadTech 1730 LCR Digibridge is an easily programmable, microprocessor-based high performance passive component tester with excellent low impedance measurement and fast measurement speeds even at 100/120Hz. It measures 12 different parameters with a basic accuracy of 0.1%. Programmable test frequency, speed, mode and voltage permits great versatility in component and materials testing. Standard IEEE-488 and Handler interfaces provide data transfer and component automation applications for faster production line testing.

Description

12 Measurement Parameters: Measure and display two of 12 impedance parameters simultaneously with a basic accuracy of 0.1%. Test results can also be shown as Deviation or %Deviation from nominal value.

7 Test Frequencies: User selectable: 100 Hz, 120 Hz, 1 kHz, 10 kHz, 20 kHz, 50 kHz and 100 kHz.

LCD Display: Large, high resolution graphic display for clear, concise results and user friendly menu interface for easy programming.

Setup Store/Recall: Store/Recall up to 50 test setups from instrument memory. Lock out front panel to ensure procedures are run the same way every time.

Programmable Source Impedance: Program tester impedance to 100., 100/25., 25. or 10. constant current when comparing results with measurements made on other testers. Measurement results can vary significantly based solely on the source impedance of the tester being used.

Voltage/Current Monitoring: Use the test signal monitoring function to display the actual voltage level across and current through the DUT.

DC Bias: External and internal DC Bias voltage from 0-5V, permits capacitors to be tested under real DC bias conditions.

Open/Short Zeroing: Reduce the effects of error sources (stray admittance/residual impedance) between the DUT and the calibrated connection to the measuring instrument.

Fast Testing: Select from one of three measurement speeds: slow, medium or fast for varying degrees of accuracy and speed up to 62 measurements/second.

Interfaces Standard: The 1730 unit includes IEEE-488 and Handler interfaces for adaption into production environments or remote control applications.

FEATURES

Parameter	*Measurement Range*
Ls, Lp	0.001 uH–99.999 kH
Cs, Cp	0.001 pF–9.9999 F
Rs, Rp	0.01 m–99.99 M.
D	0.0001–9999
Q	0.0001–9999
ESR	0.01 m–99.99 M.
LZl	0.01 m–99.99 M.
Xs	0.01 m–99.99 M.
Phase	–180.00°–+180.00°

Accuracy:	Basic LCR ±0.1%. Basic DQ ±0.001
Test Frequencies:	100 Hz, 120 Hz, 1 kHz, 10 kHz, 20 kHz, 50 kHz and 100 kHz Accuracy ±(0.01% ± 0.01Hz)
Measurement Speed:	Fast: 62 measurements/second Medium: 19 measurements/second Slow: 3 measurements/second
Ranging:	Automatic or User Selectable
Trigger:	Internal External (IEEE or Handler) Manual
Source Impedance:	100/25., 100., 25. or 10.c.c. (constant current)
AC Test Signal:	Voltage: 10 mV–1V, 10 mV steps
DC Bias Voltage:	External: 0–5 V
Internal:	0-5 V with RBIAS external resistor
DC Bias Current:	Internal: 0–200 mA using RBIAS external resistor
Display:	240 by 64 LCD display

Results Format:	Value, Deviation, % Deviation, Bin Number, GO/NO GO
Measurement Delay:	Programmable from 0-10 s in 1ms steps
Averaging:	Programmable from 1-256
Binning:	Eight pass bins for LCR One fail bin, LCR and DQ
Interfaces:	IEEE-488 and Handler Standard
Equivalent Circuit:	Series or Parallel
Zeroing:	Open and short circuit compensation.
Buzzer:	Programmable: HI LO or OFF for Pass or Fail Result
Comparator:	Set Upper/ Lower Limit as absolute value or % value for primary and secondary parameters
Test Signal:	Monitor voltage across and current through device under test.
Test Terminals:	Front: 4 BNC & 1 GND Rear: EXT BIAS: 1 BNC, 2-terminal strip
Setup Storage:	Store/Recall 50 test setups
Lock:	Keypad Lockout
Dimensions:	(w × h × d): 12.5 × 4.0 × 13.5 in (312.5 × 100.0 × 337.5 mm)
Weight:	12.4 lbs. (5.7 kg) net, 17 lbs. (7.8 kg) ship
Environmental:	Specifications: +15°C to +35°C, 75% RH Operating: 0°C to +50°C
Storage:	−40°C to +75°C
Humidity:	<85%
Power:	• 90–125 V AC, • 190 – 250 V AC, • 50 or 60 Hz, • 65W max

FUNCTION GENERATOR/ARBITRARY WAVEFORM GENERATOR

Waveform Specifications

Standard: Sine, square, triangle, ramp, noise, sin (x)/x exponential rise and fall, cardiac, dc volts

Arbitrary:	Waveform length	8 to 16,000 points
	Amplitude resolution	12 bits
	Sample rate	40 MSa/s
	Nonvolatile memory	Four waveforms (each from 8 to 16,000 points)

Frequency Specifications

Sine	100 μHz–15 MHz
Square	100 μHz–15 MHz
triangle	100 μHz–100 kHz
Ramp	100 μHz–100 kHz
White noise	10 MHz bandwidth
Resolution	10 μHz or 10 digits
Accuracy	10 ppm in 90 days (18–28°C)

Sinewave Spectral Purity Specifications

Harmonic distortion	dc to 20 kHz	–70 dBc
	20 kHz to 100 kHz	–60 dBc
	100 kHz to 1 MHz	–45 dBc
	1 MHz to 15 MHz	–35 dBc
Total Harmonic distortion	dc to 20 kHz	< 0.04%

Output Characteristics

Amplitude (into 50 ohm)	50 mVp-p to 10 Vp-p	
Accuracy (at 1 kHz)	±1% of specified output	
Flatness (sinewave relative to 1 kHz)	< 100 kHz	±1% (0.1 dB)
	100 kHz to 1 MHz	±1.5% (0.15 dB)
	1 MHz to 15 MHz	±2% (0.2 dB)

Modulation

AM

Carrier –3 dB Frequency	10 MHz (typical)
Modulation	Any internal waveform including Arb
Frequency	10 mHz–20 kHz
Depth	0%–120%
Source	Internal/External

FM

Modulation	Any internal waveform including Arb
Frequency	10 mHz–10 kHz
Deviation	10 mHz–15 MHz
Source	Internal Only

FSK

Internal rate	10 mHz–50 kHz
Frequency range	10 mHz–15 MHz
Source	Internal/External (1 MHz max.)

Burst

Carrier frequency	5 MHz max.
Count	1 to 50,000 cycles
Start phase	–360° to +360°
Internal rate	10 mHz–50 kHz ±1%
Gate source	Internal/External Gate
Trigger	Single, external or internal rate

Option 001 Phase Lock/TCXO Timebase

Timebase accuracy

Stability	±1 ppm 0°C –50°C
Aging	< 2 ppm in first 30 days (continuous operation)
	0.1 pm/month (after first 30 days)

External reference/Input	
Lock range	10 MHz ± 50 Hz

Internal reference/Output	
Frequency	10 MHz

Environmental Specifications	
Power	110 V/120 V/220 V/240 V ± 10%
Net weight	4 kg (8.8 lbs)
Size (W × H × D)	254.4 × 103.6 × 374 mm (10.0 × 4.0 × 15.1 in)
Warranty	3 Years

LOGIC ANALYZER

Features

- MagniVu$^{\text{TM}}$ acquisition technology provides 500 ps timing resolution on all channels all the time through a single probe
- Up to 200 MHz state acquisition with 400 MHz data rate supports advanced processors and buses
- Simultaneous state and high speed timing analysis through the same probes pinpoints elusive integration faults
- Broad selection of processor and bus support
- Up to 16 Mbits per channel with hardware-accelerated waveform display and search functions to rapidly analyze large amounts of data to find elusive problems far back in time
- 64 channel pattern generator with up to 268 MHz and up to 2 MB memory depth provides digital stimulus for functional verification, debugging and stress testing
- Four channel digitizing oscilloscope with up to 1 GHz, 5 GHz provides high-fidelity signal quality measurements of digital signals
- Expansion mainframe supports up to 16 modules with up to 2,176 LA channels, 1,024 pattern generator channels or 64 digitizing oscilloscope channels for large, multiple processor and bus applications
- Universal HLL source code support for correlating high-level language source with real-time trace
- Performance analysis support for optimizing target system performance
- Remote control using microsoft COM/DCOM technology supports advanced data analysis
- Microsoft® Windows®-based PC platform provides familiar user interface with network connectivity
- All modules, probes, software and accessories fully interchangeable between portable and benchtop mainframes with flexibility for future upgrades to protect your investment

Applications

- Hardware debug and verification
- Processor/bus debug and verification
- Embedded software Integration, debug and verification
- Embedded real-time software optimization
- System validation

UNIVERSAL COUNTER

TIME BASE

Internal Time Base Stability

	Standard (0° to 50°C)	*Medium oven* (Option 001)	*High oven* (Option 010)
Temperature Stability (referenced to 25°C)	$< 5 \times 10^{-6}$	$< 2 \times 10^{-7}$	$< 2.5 \times 10^{-9}$
Aging rate (after 30 days)	Per month: $<3 \times 10^{-7}$	Per day: $< 4 \times 10^{-8}$ Per month: $< 2 \times 10^{-7}$	Per day: $< 5 \times 10^{-10}$ Per month: $< 1.5 \times 10^{-8}$
Turn-on Stability vs. Time (in 30 minutes)		$< 2 \times 10^{-7}$ referenced to 2 hr	$< 5 \times 10^{-9}$ referenced to 24 hr
Calibration	Manual Adjust	Electronic	Electronic

Note that power to the time base is maintained when the counter is placed in standby via the front panel switch. The internal fan will continue to operate when in standby to maintain long-term measurement reliability.

INSTRUMENT INPUTS

Input Specifications

Channel 1 & 21

Frequency Range	**dc coupled:** dc to 225 MHz **ac coupled:** 1 MHz to 225 MHz (50 W), 30 Hz to 225 MHz (1 MW)
FM Tolerance	25%
Voltage Range and Sensitivity (Sinusoid)2	**dc to 100 MHz:** 20 mV_{rms} to ± 5 V ac + dc **100 MHz to 200 MHz:** 30 mVrms to ± 5 V ac + dc **200 MHz to 225 MHz**: 40 rms to ± 5 V ac + dc (all specified at 75 mVrms with opt. rear connectors)[3]
Voltage Range and Sensitivity (Single-Shot Pulse)2	**4.5 ns to 10 ns pulse width:** 100 mVpp to 10 Vpp (150 mVpp with optional rear connectors)[3] **10 ns pulse width:** 50 mVpp to 10 Vpp (100 mVpp with optional rear connectors)[3]

Trigger Level2	
Range	± 5.125 V
Accuracy	± (15 mV + 1% of trigger level)
Resolution	5 mV
Damage Level	
50 W	5 V_{rms}
0 to 3.5 kHz, 1 MW	350 Vdc + ac pk
3.5 kHz to 100 kHz, 1 MW	350 Vdc + ac pk linearly derated to 5 V_{rms}
>100 kHz, 1 MW	5 V_{rms}
Input Characteristics	
Channel 1 & 21	
Impedance	1 MW to 50 W
1 MW Capacitance	30 pF
Coupling	ac or dc
Low-Pass Filter	100 kHz, switchable –20 dB at > 1 MHz
Input Sensitivity	Selectable between Low, Medium, or High (default). Low is approximately 2 × High Sensitivity.
Trigger Slope	Positive or Negative
Auto Trigger Level	
Range	0 to 100% in 10% steps
Frequency	> 100 Hz
Input Amplitude	> 100 mVpp (No amplitude modulation)
Attenuator	
Voltage Range	×10
Trigger Range	×10
Input Specifications	
Channel 3	
Frequency Range	
Option 030	100 MHz to 3 GHz
Option 050	200 MHz to 5 GHz
Option 124	200 MHz to 12.4 GHz
Power Range and Sensitivity (Sinusoid)	
Option 030	100 MHz to 2.7 GHz: –27 dBm to +19 dBm 2.7 GHz to 3 GHz: –21 dBm to +13 dBm
Option 050	200 MHz to 5 GHz: –23 dBm to +13 dBm
Option 124	200 MHz to 12.4 GHz: –23 dBm to +13 dBm
Damage Level	
Option 030	5 Vrms
Option 050	+25 dBm
Option 124	+25 dBm
Characteristics	
Impedance	50 W
Coupling	AC
VSWR	< 2.5:1

External Arm Input Specifications

Signal Input Range	TTL compatible Pulse width: > 50 ns
Timing Restrictions	Transition time: < 250 ns Start-to-stop time: > 50 ns
Damage Level	10 Vrms

External Arm Input Characteristics

Impedance	1 kW
Input Capacitance	17 pF
Start/Stop Slope	Positive or Negative

External Time Base Input Specifications

Voltage Range	200 mVrms to 10 Vrms
Damage Level	10 Vrms
Frequency	1 MHz, 5 MHz, and 10 MHz (53132A 10 MHz only)

Time Base Output Specifications

Output Frequency	10 MHz
Voltage	> 1 Vpp into 50 W (centered around 0 V)

MEASUREMENT SPECIFICATIONS

Frequency

Channel 1 and 2	**Range:** 0.1 Hz to 225 MHz
Channel 3	**Option 030:** 100 MHz to 3 GHz **Option 050:** 200 MHz to 5 GHz **Option 124:** 200 MHz to 12.4 GHz

Period

Channel 1 and 2	**Range:** 4.44 ns to 10 s
Channel 3	**Option 030:** 0.33 ns to 10 ns **Option 050:** 0.2 ns to 5 ns **Option 124:** 80 ps to 5 ns

Frequency Ratio

Measurement is specified over the full range of each input.

Results Range	10^{-10} to 10^{11}
'Auto' Gate Time	100 ms

For Automatic or External Arming (and signals < 100 Hz using Timed Arming)

LSD Displayed	(t_{res}/Gate Time) × Frequency or Period
RMS Resolution	($\sqrt{t^2_{res} + 2 \times \text{Trigger Error}^2}$)/Gate Time) × Frequency or Period
Typical (t_{res})	650 ps
For Automatic Arming	Gate Time = N/Frequency where $N = 1$ for standard channel Frequency < 1 MHz, 4 for standard channel Frequency > 1 MHz, 128 for optional channel
Systematic Uncertainty	(± Time Base Error ± t_{acc}/Gate Time) × Frequency or Period
Typical (t_{acc})	350 ps
worst case	1.25 ns
Trigger	Default setting is Auto Trigger at 50%

For Time or Digits Arming

LSD Displayed	(2√2 × t_{res}/Gate Time × Number of Samples) + t_{jitter}/Gate Time) × Frequency or Period	
RMS Resolution	(4 × t^2_{res} + (2 × Trigger Error^2)/Gate Time × ÖNumber of Samples + t_{jitter}/Gate Time) × Frequency or Period	
	t_{res}	t_{jitter}
Typical	500 ps	50ps
Number of Samples	Gate Time × Frequency (Frequency < 200 kHz) Gate Time × 200,000 (Frequency > 200 kHz)	
Systematic Uncertainty	(± Time Base Error ± t_{acc}/GateTime) × Frequency or Period	
Typical (t_{acc})	100 ps	
Worst case	300 ps	
Trigger	Default setting is Auto Trigger at 50%	

Time Interval

Measurement is specified over the full signal ranges of Channels 1 and 2.

Results Range	–1 ns to 10^5 s
LSD	500 ps

Phase

Measurement is specified over the full signal range of Channels 1 and 2.

Results Range	–180° to +360°

Duty Cycle

Measurement is specified over the full signal range of Channel 1. However, both the positive and negative pulse widths must be greater than 4 ns.

Results Range	0 to 1 (e.g. 50% duty cycle would be displayed as .5)

Rise/Fall Time

Measurement is specified over the full signal ranges of Channel 1. The interval between the end of one edge and start of a similar edge must be greater than 4 ns.

Edge Selection	Positive or Negative

Use of the input attenuator multiplies all voltage specifications (input range, results range, resolution and systematic uncertainty) by a nominal factor of 10.

Time Interval, Pulse Width, Rise/Fall Time

RMS Resolution	$\sqrt{}(t_{res})^2$ + Start Trigger Error^2 + Stop Trigger Error^2
Trigger	Default setting is Auto Trigger at 10% and 90%
Results Range	5 ns to 10^5 ns
LSD	500 ps

Pulse Width

Measurement is specified over the full signal range of Channel 1. The width of the opposing pulse must be greater than 4 ns.

Pulse Selection	Positive or Negative
Trigger	Default setting is Auto Trigger at 50%
Results Range	5 ns to 10^5 ns
LSD	500 ps

Totalize

Measurement is specified over the full signal range of Channel 1.

Results Range	0 to 10^{15}
Resolution	± 1 count

Peak Volts

Measurement is specified on Channels 1 and 2 for dc signals; or for ac signals of frequencies between 100 Hz and 30 MHz with peak-to-peak amplitude greater than 100 mV.

Results Range	–5.1 V to +5.1 V
Resolution	10 mV
Peak Volts Systematic Uncertainty	for ac signals: 25 mV + 10% of V for dc signals: 25 mV + 2% of V
Systematic Uncertainty	± (Time Base Error × Measurement) ± Trigger Level Timing Error ± 1.5 ns Differential Channel Error
Frequency Ratio	Ch1/Ch2, Ch1/Ch3, Ch2/Ch1, Ch3/Ch1
LSD	**Ratio 1/2:** 1/Ch2 Freq × Gate Time **Ratio 2/1:** Ch2 Freq/ (Ch1 Freq)^2 × Gate Time
RMS Resolution	**Ratio 1/2:** 2 × Ö1 + (Ch1 Freq × Ch2 Trigger Error)^2/Ch2 Freq × Gate Time **Ratio 2/1:** 2 × Ch2 Freq × Ö1 + (Ch1 Freq × Ch2 Trigger Error)^2/ (Ch1 Freq)^2 × Gate Time

For measurements using Ch3, substitute Ch3 for Ch2 in these equations. To minimize relative phase measurement error, connect the higher frequency signal to channel 1.

Systematic Uncertainty	± 2 × resolution
Phase	
RMS Resolution	$\sqrt{}((t_{res})^2 + (2 \times \text{Trigger Error}^2)) \times (1 + (\text{Phase}/360°)^2) \times$ Frequency × 360°
Systematic Uncertainty	(± Trigger Level Timing Error ± 1.5 ns Differential Channel Error) × Frequency × 360°
Duty Cycle	
RMS Resolution	$\sqrt{}((t_{res})^2 + (2 \times \text{Trigger Error}^2)) \times (1 + \text{Duty Cycle}^2) \times$ Frequency t_{res} = 750 ps
Gate Time	Auto Mode, or 1 ms to 1000 s
Measurement Throughput	HP-ASCII: 200 measurements/s (maximum)
Measurement Arming	
Start Measurement	Free Run, Manual, or External
Stop Measurement	Continuous, Single, External, or Timed
Time Interval	100 μs to 10 s
Delayed Arming	100 ns to 10 s
Arming Modes	(Note that not all arming modes are available for every measurement function.)

Auto Arming	Measurements are initiated immediately and acquired as fast as possible, using a minimum of signal edges.
Timed Arming	The duration of the measurement is internally timed to a user-specified value (also known as the "gate time").
Digits Arming	Measurements are performed to the requested resolution (number of digits) through automatic selection of the acquisition time)
External Arming	An edge on the External Arm Input enables the start of each measurement. Auto Arming, Timed arming modes or another edge on the External Arm Input may be used to complete the measurement.
Time Interval Delayed Arming	For Time Interval measurements, the Stop Trigger condition is inhibited for a user-specified time following the Start Trigger.

Measurement Limits

Limit Checking	The measurement value is checked against user-specified limits at the end of each measurement.
Display Modes	The measurement result may be displayed as either the traditional numeric value or graphically as an asterisk moving between two vertical bars.
Out-of-Limits Indications	The limits annunciator will light on the front panel display. The instrument will generate an SRQ if enabled via HP-IB. The limits hardware signal provided via the RS-232 connector will go low for the duration of the out-of-limit condition. If the Analog Display mode is enabled, the asterisk appears outside the vertical bars, which define the upper and lower limits.

Fractional Time Base Error

Time base error is the maximum fractional frequency variation of the time base due to aging or fluctuations in ambient temperature or line voltage:

Time Base Error = ($\Delta f/f$ aging rate + $\Delta f/f$ temperature + $\Delta f/f$ line voltage)

Multiply this quantity by the measurement result to yield the absolute error for that measurement. Averaging measurements will not reduce (fractional) time base error. The counters exhibit negligible sensitivity to line voltage; consequently the line voltage term may be ignored.

Trigger Error

External source and input amplifier noise may advance or delay the trigger points that define the beginning and end of a measurement. The resulting timing uncertainty is a function of the slew rate of the signal and the amplitude of spurious noise spikes (relative to the input hysteresis band). The (rms) trigger error associated with a single trigger point is:

Trigger Error = $\surd(E_{\text{input}})^2 + (E_{\text{signal}})^2$ / Input Signal Slew Rate at Trigger Point (in seconds)

where

E_{input} = RMS noise of the input amplifier: 1 mVrms (350 μVrms typical). Note that the internal measurement algorithms significantly reduce the contribution of this term.

E_{signal} = RMS noise of the input signal over a 225 MHz bandwidth (100 kHz bandwidth when the low-pass filter is enabled). Note that the filter may substantially degrade the signal's slew rate at the input of the trigger comparator.

For two-trigger-point measurements (e.g. Rise Time, Pulse Width), the Trigger Errors will be referred to independently as Start Trigger Error and Stop Trigger Error.

Trigger Level Timing Error

Trigger level timing error results from a deviation of the actual trigger level from the specified level. The magnitude of this error depends on resolution and accuracy of the trigger level circuit, input amplifier fidelity, input signal slew rate, and width of the input hysteresis band.

The following equations should be summed together to obtain the overall Trigger Level Timing Error. At the "High" sensitivity input setting, the hysteresis band can be assumed to be the sensitivity of the counter input. Reduction of input sensitivity or use of the attenuator will increase the size of this band.

Input Hysteresis Error: 0.5 × Hysteresis Band/Input Signal Slew Rate at Start Trigger Point-0.5 × Hysteresis Band/Input Signal Slew Rate at Stop Trigger Point

Trigger Level Setting Error: ± 15 mV ± (1% × Start Trigger Level Setting)/Input Signal Slew Rate at Start Trigger Point: ± 15 mV ± (1% × Stop Trigger Level Setting)/Input Signal Slew Rate at Stop Trigger Point

Differential Channel Error	The differential channel error term stated in several Systematic Uncertainty equations accounts for channel-to-channel mismatch and internal noise. This error can be substantially reduced by performing a TI calibration (accessible via the Utility Menu) in the temperature environment in which future measurements will be made.

Measurement Statistics

Available Statistics	Mean, Minimum, Maximum Standard Deviation
Number of Measurements	2 to 1,000,000. Statistics may be collected on all measurements or on only those which are between the limit bands. When the Limits function is used in conjunction with Statistics, *N* (number of measurements) refers to the number of in-limit measurements. In general, measurement resolution will improve in proportion to the square root of *N*, up to the numerical processing limits of the instrument.
Measurements	Statistics may be collected for all measurements except Peak Volts and Totalize.

General Information

Save and Recall	Up to 20 complete instrument setups may be saved and recalled later. These setups are retained when power is removed from the counter.
Rack Dimensions (H × W × D)	88.5 mm × 212.6 mm × 348.mm
Weight	3.5 kg maximum
Warranty	3 years

Power Supply	**AC:** 100 to 120 VAC ± 10%—50, 60 or 400 Hz ± 10%, 220 to 240 VAC ± 10%—50 or 60 Hz ± 10% **DC (Option 002):** 10 to 32 Vdc, 4A inrush
AC Line Selection	Automatic
Power Requirements	170 VA maximum (30 W typical)
Environment	0°C to 55°C operating −40°C to 71°C storage
Remote Interface	HP-IB (IEEE 488.1–1987, IEEE 488.2–1987)
Remote Programming Language	SCPI-1992.0 (Standard Commands for Programmable Instruments)
Safety	Designed in compliance with IEC 801–2, –3, –4
EMC	CISPR-11, EN50082-1, IEC 801–2, –3, –4
Radiated Immunity Testing	When the product is operated at maximum sensitivity (20 mVrms) and tested at 3 V/m according to IEC 801–3, external 100 to 200 MHz electric fields may cause frequency miscounts.

Appendix E

Multiple Choice Questions

Chapter 1

1. A d’Arsonval galvanometer can have the instrument error due to:
   - (a) Humidity
   - (b) Friction in bearings
   - (c) High temperature
   - (d) Vibrations in the taut suspension
2. A resistance of approximate value 10 Ω is measured by the ammeter-voltmeter method with an ammeter of resistance 0.02 Ω and voltmeter of resistance 5 kΩ. The resistance should be measured
   - (a) by connecting the ammeter on the side of the unknown resistance as this connection gives better accuracy.
   - (b) by connecting the voltmeter on the side of the unknown resistance as this connection gives better accuracy.
   - (c) by any of the two connections (a) or (b), as both of them give equal accuracy.
   - (d) by none of the above methods.
3. A reverse-biased diode in addition to a forward-biased diode is used in rectifier-type voltmeter using half-wave rectification. Which of the following statement is true in respect of the reverse-biased diode?
   - (a) It does not allow any current to flow through the meter during the negative-half cycle.
   - (b) It shorts the meter during the negative-half cycle.
   - (c) It does not allow the reverse leakage current to flow through the meter during the negative-half cycle.
   - (d) All of the above.

4. In a true rms reading voltmeter two thermocouples are used in order to
   (a) increase sensitivity.
   (b) prevent drift in the dc amplifier.
   (c) cancel out nonlinear effects of the first thermocouple.
   (d) achieve all of the above.

5. Accuracy of a reading of an instrument refers to:
   (a) How close the reading is to the previous reading.
   (b) How many readings were taken at that point.
   (c) The repeatability of the reading.
   (d) How close the reading is to the actual value.

6. The principle of operation of a basic analog ohmmeter is:
   (a) A known voltage is placed across the unknown resistance and a d'Arsonval movement is used to sense the current. The d'Arsonval movement is calibrated in ohms; the scale is linear.
   (b) A known voltage is placed across the unknown resistance and a d'Arsonval movement is used to sense the current. The d'Arsonval movement is calibrated in ohms; the scale is nonlinear.
   (c) A known current is driven through the unknown resistance and a voltmeter is used to measure the voltage across the resistance. The voltmeter calibrated in ohms; the scale is linear.
   (d) A known current is driven through the unknown resistance and a voltmeter is used to measure the voltage across the resistance. The voltmeter calibrated in ohms; the scale is nonlinear.

7. The torque developed in a PMMC movement is
   (a) directly proportional to the flux density in the air gap and inversely proportional to number of turns in the coil.
   (b) directly proportional to the flux density in the air gap and inversely proportional to the effective coil area.
   (c) directly proportional to the flux density in the air gap and also directly proportional to the effective coil area.
   (d) directly proportional to the flux density in the air gap and inversely proportional to the current in the coil.

8. To decrease the response time, a PMMC movement system is designed to be
   (a) slightly overdamped.
   (b) slightly critically damped.
   (c) slightly underdamped.
   (d) highly underdamped.

9. If a 1 mA peak high frequency sinusoidal with zero dc shift is applied directly to a PMMC movement of 1 mA rating, the deflection will correspond to
   (a) 0 mA.
   (b) 0.5 mA.
   (c) 1 mA.
   (d) none of the above.

10. The spring used in a PMMC movement may weaken due to aging. If proper compensation is not adopted, this will result in
(a) increase in deflection.
(b) no change in deflection.
(c) decrease in deflection.
(d) change of flux density.

11. In an ammeter, the shunt is used to
(a) decrease the voltage range.
(b) increase the voltage range.
(c) decrease the current range.
(d) increase the current range.

12. In a voltmeter, a resistance in series with PMMC movement is used to
(a) increase the voltage range.
(b) decrease the voltage range.
(c) decrease the current range.
(d) increase the current range.

13. The sensitivity of rectifier ac meters is
(a) more than that of dc meters.
(b) less than that of dc meters.
(c) equal to that of dc meters.
(d) always infinite.

14. An unsymmetrical waveform of 5 V positive peak and 10 V negative peak is applied to a peak-to-peak reading voltmeter using a negative clamper. The meter will read
(a) 5 V
(b) –10 V
(c) 15 V
(d) –15 V

15. The scale of resistance in an ohmmeter is nonlinear because the
(a) relationship between voltage and current is nonlinear.
(b) relationship between voltage and resistance is nonlinear.
(c) relationship between resistance and current is nonlinear.
(d) PMMC movement gives nonlinear readings.

16. A voltmeter has 100 divisions on 0–100 V range. One-tenth of each division can be read with certainty. The resolution of the meter is
(a) 0.1 V
(b) 0.01 V
(c) 1.0 V
(d) 1.01 V

17. The resistance of a voltmeter in 0–100 V range is 1000 Ω. Its sensitivity is
(a) 0.1 Ω/V
(b) 1 Ω/V
(c) 10 Ω/V
(d) 100 Ω/V

18. A voltmeter has an accuracy of ±2% of full-scale range of 200 V. This meter is used to read 100 V. The actual measured voltage will fall between
(a) 98 V and 102 V
(b) 99.8 V and 100.2 V
(c) 99.6 V and 100.4 V
(d) 96 V and 104 V

19. A 200 mA meter has an accuracy of ±2%. While reading 100 mA, its accuracy will be
(a) ±2%
(b) ±1%
(c) ±4%
(d) ±5%

20. Two voltmeters having sensitivity of 75 Ω/V and 50 Ω/V are connected in series. Each voltmeter is set for the range of 400 V. If 500 V is applied across the series combination, the readings of the meters will be respectively
(a) 300 V and 200 V
(b) 400 V and 100 V
(c) 250 V and 250 V
(d) 350 V and 150 V

Chapter 2

1. A 4-1/2 digits digital voltmeter having 100% over range capability will display 200 mV signal on 200 mV range as
(a) 200.00 mV
(b) 400.00 mV
(c) 199.99 mV
(d) 399.99 mV

2. In a 4-1/2 digits digital multimeter, used in the laboratory, on a 20 V full-scale range, the resolution of the instrument will be
(a) 1 mV
(b) 10 mV
(c) 0.1 mV
(d) 0.01 mV

3. The crest factor of a DMM specifies the
(a) maximum dc value of the signal which the DMM can handle without overloading.
(b) maximum peak value of the signal which the DMM can handle without overloading.
(c) maximum temperature at which a DMM can operate without overloading.
(d) maximum humidity at which a DMM can operate without overloading.

4. One of the factors of using a R-2R ladder D/A converter is that
(a) it requires multiple resistors of a single value.
(b) its accuracy is poorer than that of a weighted-resistor D/A converter.
(c) it needs higher value of resistors.
(d) this approach gives larger size of resistors.

5. The dynamic range of an 8-bit D/A converter is approximately
(a) 24 dB
(b) 16 dB
(c) 32 dB
(d) 48 dB

6. One of the factors in favour of using a dual slope A/D converter is that the
(a) output generated is independent of the component values of the integrator.
(b) output generated is dependent on the component values of the integrator.
(c) output generated is directly proportional to the reference voltage.
(d) conversion time is inversely proportional to the number of bits of the counter.

7. For an N-bit successive approximation A/D converter, the conversion time
(a) is independent of N.
(b) cannot be estimated.
(c) is equal to $1/Nf_{\text{CLK}}$.
(d) is equal to N/f_{CLK}.

8. In undersampling, A/D converters are used at the sampling rate that is
(a) higher than the Nyquist rate.
(b) lower than the Nyquist rate.
(c) independent of the Nyquist rate.
(d) equal to the Nyquist rate.

9. The value of the capacitor in a sample-and-hold circuit should be
(a) small enough to minimize droop.
(b) large enough to minimize droop.
(c) large enough to track faster signals.
(d) of any value.

10. A 4-bit flash A/D converter needs
(a) 4 analog comparators.
(b) 8 analog comparators.
(c) 7 analog comparators.
(d) 15 analog comparators.

11. In a sigma delta A/D converter, the signal-to-noise ratio
(a) reduces with over sampling ratio.
(b) is independent of the over sampling ratio.
(c) increases with over sampling ratio.
(d) is much higher than other types of A/D converters.

12. The count equivalent of a 4-4/5 Digit DMM is
(a) 4000
(b) 5000
(c) 40000
(d) 50000

13. The CMRR of a DMM for a frequency of 50 Hz is given as 100 dB. The measured voltage for an input difference voltage of 1 mV is
(a) 5 mV
(b) 100 V
(c) 10 V
(d) 1 V

14. A 2-wire remote sensing method used for the measurement of an unknown resistance using a DMM introduces an error of
(a) $2R_{\text{lead}}$
(b) R_{lead}
(c) $-2R_{\text{lead}}$
(d) $-R_{\text{lead}}$

15. The value of a capacitor is measured using a DMM by passing a constant current of 1 mA through the capacitor. If the voltage measured at 0.1 ms is 1 V, then the value of the unknown capacitor is
(a) 1 F
(b) 1 μF
(c) 0.1 μF
(d) 0.01 μF

Chapter 3

1. Triggering in a CRO
(a) generates the sweep signal.
(b) provides input to the vertical plates.
(c) provides stability in a repeated waveform.
(d) chops the input signal.

2. Z modulation in a CRO display
(a) generates a signal in Z-axis.
(b) provides the third dimension of intensity to the traditional waveform display.
(c) provides depth of the waveform.
(d) indicates depth of the CRT.

3. Basic function of a trigger in a CRO is to
(a) increase the bandwidth.
(b) stabilize repeating waveforms.
(c) adjust the amplitude.
(d) adjust the phase.

4. The colour of the beam in a CRO is the characteristic of the
(a) signal being viewed.
(b) primary electrons being emitted from the cathode.
(c) coating material of the screen.
(d) acceleration voltage.

5. If the retrace is visible on the CRT, the problem may be due to
   (a) loss in the sync signal.
   (b) intensity being too high.
   (c) blanking control being faulty.
   (d) horizontal gain being too high.

6. The time base of a CRO is developed by a
   (a) sine waveform.
   (b) square waveform.
   (c) rectangular waveform.
   (d) sawtooth waveform.

7. The input impedance of a general-purpose CRO must
   (a) be high.
   (b) be low.
   (c) appear capacitive.
   (d) appear inductive.

8. The delay line is used in an analog CRO to introduce delay in the
   (a) vertical path to compensate for the delay in the horizontal path.
   (b) horizontal path to compensate for the delay in the vertical path.
   (c) sweep generator.
   (d) triggering.

9. If the sweep generated is not linear, the beam
   (a) will move quickly.
   (b) will move slowly.
   (c) will move quickly for some time and slowly for some time.
   (d) movement will not be affected.

10. The deflection on the screen will be more if the
   (a) deflection voltage is decreased.
   (b) accelerating voltage is increased.
   (c) distance between the deflection plates is increased.
   (d) distance from the centre of the deflection plates to screen is increased.

11. The resolution of 1024 × 768 of a raster display device means that there are
   (a) 1024 rows (scan-lines).
   (b) 768 pixels in each scan-line.
   (c) 1024 pixels in each scan-line.
   (d) 1024 rows and 768 pixels in each row.

12. In an equivalent-time sampling method
   (a) the picture is acquired in successive cycles of the waveform and then constructed from these samples.
   (b) a few sample points of the signal are collected in a single pass and linear interpolation is used to fill in the gaps.
   (c) a few sample points of the signal are collected in a single pass and sin x/x interpolation is used to fill in the gaps.
   (d) only digital waveforms can be sampled.

13. In a DPO,
(a) intensity modulation cannot be provided.
(b) intensity modulation can be provided.
(c) storage facility is not available.
(d) horizontal input is not available.

14. A sinusoidal signal with a dc shift can be displayed by using
(a) ground input coupling.
(b) ac input coupling.
(c) dc input coupling.
(d) high bandwidth CRO.

15. The high frequency noise signal riding on a small frequency signal can be seen on a CRO by
(a) adjusting the vertical gain of the CRO.
(b) using delayed triggering.
(c) adjusting the horizontal gain of the CRO.
(d) adjusting intensity.

16. Lissajous patterns can be used to determine
(a) amplitude of the signal.
(b) amplitude distortion.
(c) only the phase shift between the two signals.
(d) frequency and phase relationship between two signals.

17. To measure rise time of greater than 3.5 ns, we should choose a CRO with
(a) maximum bandwidth of 100 MHz.
(b) minimum bandwidth of 100 MHz.
(c) minimum bandwidth of 1 GHz.
(d) maximum horizontal gain of 100.

18. A sinusoidal signal of 10 kHz is applied to the vertical input of an analog CRO. The display shows three cycles of the waveform. The setting of the time base is
(a) 0.3 s
(b) 3 s
(c) 0.1 s
(d) 1 s

Chapter 4

1. The impedance of a passive probe
(a) increases with frequency.
(b) is independent of frequency.
(c) is not important.
(d) decreases with frequency.

2. If R_1 and C_1 are the resistance and capacitance of the CRO probe, and R_2 and C_2 are the resistance and capacitance of the CRO, then the probe is compensated if

(a) $R_1C_1 = R_2C_2$
(b) $R_1C_1 < R_2C_2$
(c) $R_1C_1 > R_2C_2$
(d) R_1C_1 is not related to R_2C_2

3. A 50 Ω resistive divider probe is to be designed to provide an attenuation of 10 and 20 when connected to a 50 Ω input CRO through a cable of 50 Ω. The resistance of the probe to be included in the probe tip respectively will be equal to
(a) 500 Ω and 1000 Ω
(b) 100 Ω and 500 Ω
(c) 450 Ω and 950 Ω
(d) 950 Ω and 450 Ω

4. 1000 volts dc voltage is to be measured with a CRO having input resistance of 1 MΩ and maximum dc input voltage of 10 volts. The probe tip resistance of this high voltage passive probe should be
(a) 10 MΩ
(b) 1 MΩ
(c) 100 MΩ
(d) 99 MΩ

5. The FET input active probes provide
(a) low input resistance and low input capacitance.
(b) high input resistance and low input capacitance.
(c) low input resistance and high input capacitance.
(d) high input resistance and high input capacitance.

6. A transformer-based current probe can measure
(a) only ac current.
(b) only dc current.
(c) both dc and ac currents.
(d) none of these.

7. A Hall Effect based current probe can measure
(a) only ac current.
(b) only dc current.
(c) both dc and ac currents.
(d) none of these.

8. When the amp-second product is exceeded for a current probe,
(a) the relationship between current input and voltage output of the probe becomes linear.
(b) clipping of the waveform peaks is removed.
(c) the core material of the probe's coil goes into saturation.
(d) the performance of the probe does not get changed.

9. A probe can introduce
(a) inductive loading.
(b) capacitive loading.

(c) resistive loading.
(d) all of the above.

10. The capacitive loading of a probe can
(a) decrease the rise time.
(b) increase the bandwidth.
(c) increase the rise time.
(d) increase the impedance.

11. The frequency of ringing introduced by the probe is
(a) directly proportional to the ground lead inductance.
(b) inversely proportional to the square root of the ground lead inductance.
(c) directly proportional to the capacitance of the probe.
(d) directly proportional to the square root of the capacitance of the probe.

12. The longer current pulses can be represented without significant decay, when the decay time constant (L/R ratio) of a current probe
(a) is small.
(b) is large.
(c) is anything.
(d) is zero.

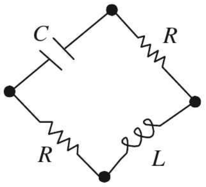
13. The maximum non-destructive input voltage (for high frequency probes) at high frequencies
(a) decreases.
(b) increases.
(c) does not change.
(d) is none of these.

14. A 100 cm long probe cable can generate a 5 ns signal delay. For a signal frequency of 1 MHz, this delay causes a phase shift of
(a) 0°
(b) 1°
(c) 1.8°
(d) 10°

15. The maximum voltage rating of a probe is determined by the
(a) breakdown voltage of the probe body.
(b) breakdown voltage of the probe components at the measuring point.
(c) breakdown voltage of the probe body or breakdown voltage of the probe components at the measuring point.
(d) input resistance of the source.

Chapter 5

1. Frequency can be measured by using
   (a) Maxwell bridge.
   (b) Schering brdge.
   (c) Heaviside Campbell bridge.
   (d) Wien bridge.
2. A Wheastone bridge cannot be used for precision measurements because errors are introduced on account of
   (a) resistance of connecting leads.
   (b) thermoelectric emf.
   (c) contact resistances.
   (d) all of the above.
3. The bridge shown
   (a) cannot be balanced.
   (b) can be balanced but the frequency of excitation must be known.
   (c) can be balanced for only one frequency.
   (d) can be balanced.



4. What type of bridge is recommended for measuring resistances less than 1 ohm?
   (a) Wheatstone bridge.
   (b) Schering bridge.
   (c) Maxwell bridge.
   (d) Kelvin bridge.
5. A 4-wire measurement method is used
   (a) to remove the effect of wire resistance.
   (b) where the source is a voltage signal.
   (c) where the input resistance of the measuring circuit is very low.
   (d) to compensate the effect of source capacitance.
6. The disadvantage of Maxwell's bridge is that
   (a) inductance measurement is frequency dependent.
   (b) it cannot be used to measure inductance for large values of Q.
   (c) both inductance and Q cannot be measured.
   (d) resistance of the inductor measurement is frequency dependent.
7. In a constant voltage Wheatstone bridge, as the value of $n = \frac{R_2}{R_4} = \frac{R_1}{R_3}$ increases,
   (a) the linearity between Thevenin voltage and change in unknown resistance decreases.
   (b) the sensitivity of the bridge increases.
   (c) the linearity between Thevenin voltage and change in unknown resistance increases and the sensitivity of the bridge decreases.
   (d) linearity and sensitivity do not change.

8. The guarded Wheatstone bridge
(a) eliminates the problem of leakage current.
(b) guards the bridge from radiations.
(c) changes the constant voltage bridge to constant current bridge.
(d) uses ac meter for detecting unbalance.

9. A 3-wire remotely connected Wheatstone bridge is used to
(a) remove the effect of wire resistance in a single-element varying bridge.
(b) remove the effect of wire resistance in an all-elements varying bridge.
(c) compensate the effect of source resistance.
(d) compensate the effect of meter capacitance.

10. The relationship between series inductance, L_s, and parallel inductance, L_p, is

(a) $L_s = L_p \dfrac{Q^2}{1+Q^2}$

(b) $L_p = L_s \dfrac{Q^2}{1+Q^2}$

(c) $L_s = L_p \dfrac{1+Q^2}{Q^2}$

(d) $L_p = L_s \dfrac{1+Q^2}{Q^2}$

11. The quality factor, Q, and dissipation factor, D
(a) are same
(b) are not related
(c) are related by $Q = 1/D$
(d) are related by $Q = 1/2D$

12. For balancing an ac bridge
(a) the product of the magnitudes of the impedances/admittances of the opposite arms must be different.
(b) the sum of the phase angles of the opposite arms must be different.
(c) the product of the magnitudes of the impedances/admittances of the opposite arms must be equal and the sum of the phase angles of the opposite arms must be different.
(d) the product of the magnitudes of the impedances/admittances of the opposite arms must be equal and the sum of the phase angles of the opposite arms must be equal.

13. The disadvantage of Hay's bridge is that
(a) the inductance measurement is frequency dependent for low values of Q.
(b) it cannot be used to measure an ideal inductor value only.
(c) both inductance and Q cannot be measured.
(d) resistance of the inductor measurement is independent of frequency for low values of Q.

14. In an ac bridge, the sum of the phase angles of the impedances of one of the opposite arms (OA1) is zero and the sum of the phase angles of the other opposite arms (OA2) is −45°. The bridge is
(a) balanced.
(b) not balanced and can be balanced by adding a capacitive element in one of the arms OA2.
(c) not balanced and can be balanced by adding an inductive element in one of the arms OA2.
(d) cannot be balanced.

15. To measure permittivity of an insulator sample, we can use
(a) Kelvin bridge.
(b) Maxwell bridge.
(c) Hay'bridge.
(d) Schering bridge.

16. Wagner Ground connection is used to
(a) eliminate stray capacitances effect.
(b) eliminate insulation resistance effect.
(c) improve low frequency response.
(d) improve dc response.

17. In the Anderson loop topology, the output is
(a) inversely proportional to the change in impedance.
(b) a linear function of the change in impedance.
(c) inversely proportional to the square of the change in impedance.
(d) directly proportional to the square of the change in impedance.

18. In an LCR bridge
(a) open/short compensation is used to compensate for stray admittance and residual impedance of the test cables.
(b) averaging is used to improve the measurement accuracy.
(c) range holding is used where repetitive testing of similar value components is to be done.
(d) all of above are used.

Chapter 6

1. A strip chart recorder records
(a) the parameters on the digital display.
(b) one or more dependent parameters that vary with reference to some independent parameter.
(c) a parameter by stripping part of the parameter.
(d) one or more parameters that vary with time.

2. Curvilinear recording mechanism in a PMMC galvanometer recorder can be changed to rectilinear recording mechanism by

(a) reducing the length of the pen.
(b) using the pivoted pen motor assembly.
(c) using the rectangular enclosure.
(d) removing part of the recorded curve.

3. As compared to other recorders, the frequency response of the CRT-based optical recorder is
(a) the highest.
(b) the lowest.
(c) lower than the ink-filled based recorders.
(d) lower than the other optical recorders.

4. The best bandwidth of a galvanometric recorder is achieved with damping factor equal to
(a) 10
(b) 5
(c) 0.1
(d) 0.7

5. The damping factor of a galvanometric recorder is decreased by
(a) decreasing the spring stiffness.
(b) adding a resistance in series with source and recorder.
(c) adding a resistance in parallel with the recorder.
(d) increasing the number of turns of wire in the coil.

6. The *X*-*Y* recorder records
(a) the parameters on digital display.
(b) one parameter with reference to the other parameter.
(c) one parameter on the *X*-axis and the other parameter as time on the *Y*-axis.
(d) one parameter on the *Y*-axis and the other parameter as time on the *X*-axis.

7. The ultrasonic pen positioning system
(a) eliminates the sliding mechanical contacts in a servo recorder.
(b) uses an infrared source.
(c) limits the frequency response of the recorder.
(d) uses an ultra violet source.

8. The chart speed of a recorder is set to 50 mm/second. If the frequency of the signal is 10 cycles/second, one cycle of the signal extends over
(a) 0.2 mm
(b) 500 mm
(c) 5 mm
(d) 0.5 mm

9. Multipoint recorders
(a) record single parameter multiple times.
(b) record multiple parameters using one pen for each input parameter.
(c) record multiple parameters using a single pen for all input parameters.
(d) record none of the above.

10. A modern paperless recorder can be
   (a) a circular chart display.
   (b) an alarm display.
   (c) a graphics display.
   (d) all of the above.

Chapter 7

1. In signal generators, energy is
   (a) created.
   (b) generated.
   (c) converted from a dc source into ac energy at a particular frequency.
   (d) supplied by ac input to the generator.

2. A sinusoidal signal generator uses a crystal oscillator. The frequency of the crystal oscillator, f_{XTAL}, is related with the series, f_s, and anti-resonance, f_a, frequencies as follows:
   (a) $f_{\text{XTAL}} > f_s$ and f_a
   (b) $f_{\text{XTAL}} < f_s$ and f_a
   (c) $f_{\text{XTAL}} = f_s$ and f_a
   (d) $f_s < f_{\text{XTAL}} < f_a$

3. To design a single pulse generator, we can use
   (a) astable multivibrator.
   (b) bistable multivibrator.
   (c) monotable multivibrator.
   (d) none of the above.

4. Phase-locked loops are used in the frequency synthesizers. Such synthesizers can be used to generate
   (a) only integer multiples of the reference frequency.
   (b) only fraction multiples of the reference frequency.
   (c) both integer and fraction multiples of the reference frequency.
   (d) frequency, independent of the reference frequency.

5. Waveform sequencing is used in the arbitrary waveform generators to
   (a) store a large number of waveforms in the memory.
   (b) to store a part of the waveform in the memory and then generate additional waveforms from this.
   (c) create a sequence of waveforms and store it in the memory.
   (d) store/create none of the above.

6. To get sustained oscillations
   (a) $|A\beta| \geq 1$ and $\angle A\beta = \pi n$ where $n = 1, 2, 3, \ldots$
   (b) $|A\beta| \geq 1$ and $\angle A\beta = 2\pi n$ where $n = 1, 2, 3, \ldots$
   (c) $|A\beta| < 1$ and $\angle A\beta = \pi n$ where $n = 1, 2, 3, \ldots$
   (d) $|A\beta| < 1$ and $\angle A\beta = 2\pi n$ where $n = 1, 2, 3, \ldots$

7. For frequency stability of an oscillator, the rate of change of phase with frequency
(a) should be as small as possible.
(b) should be as large as possible.
(c) should be nonlinear.
(d) is not important.

8. At frequency of oscillation, the feedback factor, β, in Wien bridge oscillator
(a) is real
(b) is imaginary
(c) is complex
(d) does not change

9. In a practical *RC* oscillator, to meet phase shift requirement, we need
(a) one or two *RC* sections.
(b) two *RC* sections.
(c) three or more *RC* sections.
(d) no *RC* section.

10. In a radio frequency signal generator, the frequency, f is proportional to
(a) $1/RC$
(b) $1/\sqrt{RC}$
(c) $1/LC$
(d) $1/\sqrt{LC}$

11. In a Schmitt trigger circuit, if V_{sat} is the saturation voltage of the opamp and β is the feedback factor, then the hysteresis voltage is given by
(a) $2\ \beta V_{\text{sat}}$
(b) βV_{sat}
(c) $2V_{\text{sat}}/\beta$
(d) V_{sat}/β

12. Astable multivibrators can be used to generate
(a) only a square wave.
(b) only a triangular wave.
(c) both square and triangular waves.
(d) a sine wave.

13. Normally, a function generator can be used to get the following outputs:
(a) Only a square wave.
(b) Only a triangular wave.
(c) Both square and triangular waves.
(d) A sine, and square and triangular waves

14. Double pulse delay is
(a) twice the pulse width.
(b) the interval between the leading edge medians of the double pulses.
(c) the interval between the leading edge medians of the two single pulses.
(d) equal to delay in single pulse.

15. The difference between the pulse width jitter and pulse period jitter is
   (a) none.
   (b) that the width jitter is the stability of a pulse with reference to the previous leading edge whereas the period jitter is the stability of the trailing edge with reference to the leading edge.
   (c) that the period jitter is the stability of a pulse with reference to the previous leading edge whereas the width jitter is the stability of the trailing edge with reference to the leading edge.
   (d) infinite.

16. A digital analog synthesizer uses a
   (a) multiplexer for frequency multiplication.
   (b) mixer basically for frequency addition.
   (c) low-pass filter to filter low frequencies generated by the mixer.
   (d) frequency multiplier.

17. A phase locked loop (PLL) consists of
   (a) only a voltage controlled oscillator.
   (b) only a phase detector.
   (c) a low-pass filter, a voltage-controlled oscillator, and a phase detector.
   (d) a frequency multiplier.

18. A multiple-loop indirect synthesizer incorporates
   (a) only a multiplier PLL.
   (b) only a summing PLL.
   (c) both multiplier and summing PLLs.
   (d) only a frequency multiplier.

19. To synthesize a frequency of 2.572 MHz from a reference frequency of 1 MHz using a programmable decade indirect synthesizers, we need
   (a) 2 numbers of PLLs.
   (b) 3 numbers of PLLs.
   (c) 4 numbers of PLLs.
   (d) a single PLL.

20. A direct digital synthesizer uses
   (a) a phase accumulator, a waveform memory (ROM), and a D/A converter.
   (b) a phase accumulator and a waveform memory (ROM).
   (c) a phase accumulator and a D/A converter.
   (d) a waveform memory (ROM) and a D/A converter.

21. In a direct digital synthesizer
   (a) frequency resolution depends on the number of output bits of the ROM.
   (b) amplitude resolution depends on the number of bits of the accumulator.
   (c) both frequency and amplitude resolutions depend on the number of output bits of the ROM.
   (d) frequency resolution depends on the number of bits of the accumulator and amplitude resolution depends on the number of output bits of the ROM.

22. An AWG is designed to generate output of 100 kHz. If the clock frequency used is 100 MHz and the AWG contains four basic cycles, the memory width is
(a) 1000 points.
(b) 4000 points.
(c) 100 points.
(d) 400 points.

23. A digital generator does not require:
(a) D/A converter
(b) Clock
(c) Memory
(d) Address generator

24. A digital generator
(a) includes only a sequencer.
(b) can generate hundreds of digital signal output channels.
(c) cannot accept data from other instruments.
(d) is basically used to analyze the amplitude changes in the signal.

Chapter 8

1. The difference between a distortion analyzer (DA) and a wave analyzer (WA) is:
(a) DA is used to measure total distortion whereas WA is used to measure individual component of harmonic distortion.
(b) DA is used to measure individual component of harmonic distortion whereas WA is used to measure total distortion.
(c) Nothing.
(d) DA is used at low frequencies and WA is used at higher frequencies.

2. The spectrum analyzer is used to display
(a) only the fundamental frequency component of the waveform.
(b) whole of the frequency spectrum of the waveform.
(c) whole of the time domain signal.
(d) only the harmonics of the waveform.

3. The difference between the spectrum analyzer (SA) and FFT analyzer (FFTA) is:
(a) SA gives information about the frequency and phase whereas FFTA gives information about the frequency only.
(b) SA gives information about the frequency whereas FFTA gives information about the frequency and phase.
(c) Nothing.
(d) SA is used at low frequencies and FFTA is used at higher frequencies.

4. The logic state analyzer displays
(a) the timing diagram of the waveforms.
(b) the frequency spectrum of the waveforms.
(c) the program flow of the system under test (SUT).
(d) both the timing and frequency spectrum of the waveforms

5. Harmonic distortion is due to
   (a) dc level of the signal.
   (b) distortion in the measuring equipment.
   (c) nonlinear device or system.
   (d) none of the above.

6. The rms value of
   (a) second harmonic component is less than that of the third harmonic component.
   (b) fourth harmonic component is more than that of the third harmonic component.
   (c) second harmonic component is more than that of the third harmonic component.
   (d) second harmonic component is always zero.

7. Intermodulation distortion
   (a) occurs due to the nonlinearity of a device or system with multiple input frequencies.
   (b) is same as harmonic distortion.
   (c) is same as signal-to-noise ratio.
   (d) cannot be measured.

8. A filter bank spectrum analyzer with large frequency range and good resolution requires
   (a) only one filter and one detector.
   (b) a large quantity of hardware.
   (c) a large number of filters but only a single detector.
   (d) a large number of detectors but only a single filter.

9. A superheterodyne spectrum analyzer
   (a) does not require any mixer.
   (b) does not require any local oscillator.
   (c) mixes the frequencies in the above audio range.
   (d) does not require any IF filter.

10. Many spectrum analyzers use log scale to display a
   (a) large range of amplitude.
   (b) small range of amplitude.
   (c) large range of frequency.
   (d) small range of frequency.

11. A low-pass filter is used in a superheterodyne spectrum analyzer
   (a) to filter dc from the signal.
   (b) after the IF stage.
   (c) to filter image frequencies.
   (d) for impedance matching.

12. Two or more mixing stages are added in a superheterodyne spectrum analyzer to
   (a) increase frequency resolution.
   (b) decrease frequency resolution.
   (c) increase amplitude resolution.
   (d) decrease amplitude resolution.

13. Phase noise in a superheterodyne spectrum analyzer can
   (a) increase frequency resolution.
   (b) decrease frequency resolution.
   (c) limit the ability of the analyzer to resolve signals of unequal amplitude.
   (d) limit the ability of the analyzer to resolve signals of unequal frequency.

14. The dynamic range of a spectrum analyzer with third order intercept point of 20 dBm and a noise level of 70 dBm is
   (a) 33.3 dB.
   (b) 13.2 dB.
   (c) 46.6 dB.
   (d) 60 dB.

15. FFT algorithms
   (a) are slower than the DFT algorithms.
   (b) are faster than the DFT algorithms.
   (c) always have the same speed as DFT algorithms.
   (d) cannot be executed using computers.

16. Maximum distortion is produced by:
   (a) Hanning Window
   (b) Rectangular Window
   (c) Hamming Window
   (d) Kaiser Window

17. The best analyzer for the vibration analysis is:
   (a) Spectrum Analyzer
   (b) Vector Analyzer
   (c) FFT Analyzer
   (d) DSO

18. A logic analyzer is used when we need
   (a) to see small voltage changes in the signals.
   (b) high time interval accuracy.
   (c) to see the relationship of many signals at a time.
   (d) to measure rise time of the signal.

19. In a logic timing analyzer, user-defined threshold is used to
   (a) cut off undesired frequencies.
   (b) decide 0 and 1 transition points for the input signal.
   (c) decide the sampling frequency.
   (d) decide the memory capacity.

20. The transition detector is used in a logic timing analyzer to
   (a) use memory more efficiently.
   (b) store input signal at all sample points.
   (c) reduce the input amplitude.
   (d) generate 0 and 1 data from the input.

21. A logic timing analyzer
  (a) starts the data capture at the trace point.
  (b) stops the data acquisition after the trace point is found.
  (c) cannot display the data before the trace point.
  (d) has the facility for setting level and slope triggering.

22. Plug-on-devices (PODs) are
  (a) active devices.
  (b) used to connect display devices to the logic analyzer.
  (c) used to magnify the display.
  (d) the probes formed by grouping a number of channels.

Chapter 9

1. In a conventional frequency counter if the counting register counts up to 1500 and the frequency displayed is 150 kHz, the selected gate time is
  (a) 1 ms
  (b) 10 ms
  (c) 1 μs
  (d) 10 μs

2. In a frequency counter, prescaling is done
  (a) to increase the frequency response by reducing the design constraints on the main gate and counting register.
  (b) to decrease the frequency response by reducing the design constraints on the main gate and counting register.
  (c) by multiplying the input signal by a constant which is > 1.
  (d) by multiplying the display input by a constant which is > 1.

3. In a conventional period measurement counter if the counting register counts up to 1234 and the period displayed is 123.4 μs, the selected gate time is
  (a) 1 ms
  (b) 0.1 ms
  (c) 1 μs
  (d) 0.1 μs

4. Period measurement provides
  (a) low resolution at low signal frequencies.
  (b) high resolution at low signal frequencies.
  (c) measurements at only high signal frequencies.
  (d) no clue to the signal frequency.

5. The input conditioning circuit of a frequency counter uses a Schmitt trigger circuit. To effect one count in the counting register, the input
  (a) may cross only the upper threshold level.
  (b) may cross only the lower threshold level.
  (c) must cross both the upper and lower threshold levels.
  (d) may not cross any of the above two levels.

6. If the input to the counter is equal to the input to the Schmitt trigger circuit, we can define peak-to-peak sensitivity as equal to
(a) upper threshold level.
(b) lower threshold level.
(c) dc level of the input.
(d) separation between the upper and lower threshold levels.

7. The dynamic range of a frequency counter is generally limited by
(a) linear response range of the input amplifier.
(b) the counting register.
(c) the display.
(d) the time base circuit.

8. The function of the trigger level control in a frequency counter is to
(a) generate the triggering signal.
(b) control the counting register.
(c) shift the hysteresis levels to enable positive or negative pulse trains to be counted.
(d) control the display.

9. Temperature-compensated crystal oscillators use
(a) cooling arrangement for the crystals.
(b) external components connected to the crystal with the opposite temperature coefficient.
(c) heating arrangement for the crystals.
(d) regulated power supply.

10. Short-term stability affecting the accuracy of the crystal oscillators is the result of
(a) noise generated in the oscillator.
(b) voltage transients.
(c) shock and vibration.
(d) all of the above.

11. For a 100 MHz signal, the main gate switching time should be
(a) greater than 10 ns.
(b) substantially less than 5 ns.
(c) greater than 100 ns.
(d) substantially greater than 5 ns.

12. ± 1 count error occurs
(a) only at high frequencies.
(b) only at low frequencies.
(c) due to synchronization problem between the input signal and the clock frequency.
(d) only for the period measurement.

13. Accuracy of the measurement in the period measurement is affected by
(a) only ±1 count error.
(b) only ±time base error.
(c) only ±trigger error.
(d) all of the above.

14. The frequency of 10 MHz is to be measured by a universal counter with total time base error of 1 ppm. Then for 1 second gate time
(a) total time base error dominates over the ±1 count error.
(b) ±1 count error dominates over the total time base error.
(c) ±1 count error is same as the total time base error.
(d) none of the above.

15. The period of 10 ms and 100 ms is to be measured by the counter with time base of 1 ppm. The absolute time base error in both the cases respectively will be
(a) 10 ms and 100 ms.
(b) 10 μs and 100 μs.
(c) 10 ns and 100 ns.
(d) 10 s and 100 s.

16. In a reciprocal electronic counter, ±1 count error in the period measurement for input frequencies less than the clock frequencies, is
(a) always smaller than a corresponding frequency measurement.
(b) always larger than a corresponding frequency measurement.
(c) always equal to a corresponding frequency measurement.
(d) always zero.

17. If the interpolator in the reciprocal electronic counter can resolve a time of 1 ns and the basic time base frequency is 100 MHz, then the
(a) counter can measure time with resolution of ±10 ns with the interpolator.
(b) counter can measure time with resolution of ±1 ns with the interpolator.
(c) counter can measure time with resolution of ±100 ns with the interpolator.
(d) counter can measure time with resolution of ±1 μs with the interpolator.

18. Arming in the reciprocal electronic counter is used for precise
(a) display capability.
(b) a time base oscillator.
(c) main gate control.
(d) amplification of the signal.

19. The time interval measurement resolution with start trigger error = 40 ps, stop trigger error = 10 ps, input frequency = 1 MHz, gate time = 1s and ±1 count error = 10 ps is
(a) 10.05 ps.
(b) 0.05 ps.
(c) 60 ps.
(d) 50 ps.

20. A counter has the following data: Aging rate < 3×10^{-7} per month, temperature effect < 5×10^{-6} (0 to 50°C) and line voltage effect < 1×10^{-7} (±10% variation). The counter has not been calibrated for one year. Assuming negligible ±1 count error, the frequency accuracy (worst case frequency change) for 10 MHz input frequency will be
(a) 86 Hz.
(b) 87 Hz.
(c) 51 Hz.
(d) 37 Hz.

Chapter 10

1. A 10 V power supply is connected to a load of 1 kΩ. The current *I* is equal to
   (a) 10 mA.
   (b) 1 mA.
   (c) 0 mA.
   (d) 100 mA.
2. The chassis of an instrument is connected to the earth ground to
   (a) get proper current in the load.
   (b) get proper voltage across the load.
   (c) eliminate leakage path.
   (d) disconnect load from the circuit.
3. To avoid ground loops, we can use
   (a) single point ground.
   (b) ground bus.
   (c) ground plane.
   (d) any of the above which is appropriate.
4. If a system has both analog and digital circuits, it is recommended to have
   (a) separate ground for them.
   (b) common ground for them.
   (c) separate or common ground for them.
   (d) none of the above.
5. Multilayer PCBs generally use
   (a) multiple ground planes.
   (b) multiple power supply planes.
   (c) multiple ground and power supply planes.
   (d) none of the above.
6. The twisted pair cable consisting of signal and return lines eliminates
   (a) the loop area leading to inductive coupling.
   (b) the mutual inductive coupling.
   (c) both of the above.
   (d) none of the above.
7. Shielding is used to block
   (a) electrostatic fields.
   (b) magnetic fields.
   (c) EMI.
   (d) all of the above.
8. The coaxial cable
   (a) has low loss.
   (b) has low variance in characteristic impedance.
   (c) is preferred at high frequencies.
   (d) has all of the above.

9. Electrostatic discharge takes place at
   (a) very high voltage and very low current.
   (b) very low voltage and very low current.
   (c) very high voltage and very high current.
   (d) very low voltage and very high current.

10. We can protect the circuits from ESD by
   (a) grounding.
   (b) shielding.
   (c) transient limiters.
   (d) any of the above as appropriate.

11. Electrostatic voltage developed by different materials is
   (a) same and is independent of relative humidity.
   (b) same and is dependent upon relative humidity.
   (c) different and is independent of relative humidity.
   (d) different and is dependent upon relative humidity.

Chapter 11

1. The sequence of phases of a product life cycle is
   (a) need/opportunity, product design, concept development, production, development, support, upgrades, disposal.
   (b) need/opportunity, concept development, product design, production, development, support, upgrades, disposal.
   (c) concept development, need/opportunity, product design, production, development, support, upgrades, disposal.
   (d) need/opportunity, concept development, product design, development, production, support, upgrades, disposal.

2. Product design may follow
   (a) top-down or bottom-up approach.
   (b) outside-in or inside-out approach.
   (c) hybrid approach.
   (d) any of the above.

3. General issues in the circuit design include
   (a) selection of technology.
   (b) reliability.
   (c) high speed design.
   (d) all of the above.

4. Starting from zero time, the failure rates sequence is
   (a) intrinsic failure period, early failure period, wearout failure period.
   (b) early failure period, intrinsic failure period, wearout failure period.
   (c) wearout failure period, early failure period, intrinsic failure period.
   (d) early failure period, wearout failure period, intrinsic failure period.

5. Steady-state availability of a system is given by:

(a) $\lim_{t \to \infty} A(t) = \dfrac{\text{MTTR}}{\text{MTTR} + \text{MTTF}}$

(b) $\lim_{t \to \infty} A(t) = \dfrac{\text{MTTF}}{\text{MTTR} + \text{MTTF}}$

(c) $\lim_{t \to \infty} A(t) = \dfrac{\text{MTTR}}{\text{MTTR} + \text{MTBF}}$

(d) $\lim_{t \to \infty} A(t) = \dfrac{\text{MTTR}}{\text{MTTR} + \text{MDT}}$

6. High speed design should consider
(a) bandwidth.
(b) decoupling.
(c) crosstalk.
(d) all of the above.

7. Most of the low-power consumption instruments use CMOS devices. If f is the frequency, C the load capacitance and V the dc power supply voltage, then the power consumed is given by
(a) $P = fCV$
(b) $P = fCV^2$
(c) $P = f/CV^2$
(d) $P = fC/V^2$

8. Proper routing of the signal traces (tracks)
(a) increases trace density.
(b) helps in minimizing the impedance.
(c) reduces the effects of noise, crosstalk, EMI, etc.
(d) effects all of the above.

9. Which of the following should not be used for effective shielding?
(a) Placing decoupling capacitors near the IC packages.
(b) Slots in the power and return planes.
(c) Routing analog traces over analog return.
(d) Routing digital traces over digital return.

10. Which of the following guidelines is not appropriate for cabling?
(a) Permit easy inspection and testing.
(b) Prevent strain on conductors, connectors, and terminals.
(c) Use square holes for penetration.
(d) Allow reasonable bend radius.

11. Enclosures for electronic instruments must consider the factors such as
(a) temperature, cooling, humidity, and condensation.
(b) vibration and shock.
(c) EMI, and radiation.
(d) all of the above.

12. Cooling of the equipment can be achieved by
   (a) passive cooling.
   (b) forced cooling.
   (c) active cooling.
   (d) any or a combination of the above.

13. Which of the following do not characterize the aesthetics of a product?
   (a) Safety.
   (b) Finish.
   (c) Colour and tone.
   (d) Shape and features.

14. Ergonomics focuses on
   (a) proper visual displays.
   (b) compatible workplace layout.
   (c) safety.
   (d) all of the above.

15. Integrated testing includes
   (a) thermal tests, vibrations and shock tests, and EMI/RFI tests.
   (b) vibrations and shock tests only.
   (c) EMI/RFI tests only.
   (d) thermal tests only.

Chapter 12

1. To meet the necessary accuracy and reliability requirements of a product, the calibration should be ensured during
   (a) design.
   (b) manufacture.
   (c) testing.
   (d) all of the above.

2. Which of the following is not the basic element of a calibration programme?
   (a) Documenting, testing, and calibration procedures.
   (b) Adjustments to the equipment after test results.
   (c) Identification of measurement equipment that is deemed not to require calibration.
   (d) Maintenance of calibration records.

3. Which of the following is not true for the calibration interval?
   (a) It is fixed.
   (b) It can be shortened.
   (c) It can be lengthened.
   (d) It depends upon the results of the previous calibrations.

4. Starting from the top, the hierarchy of traceability chain is
   (a) working standards, transfer standards, primary standards.
   (b) primary standards, transfer standards, working standards.

(c) primary standards, working standards, transfer standards.
(d) transfer standards, primary standards, working standards.

5. If the accuracy of measurement at user level (Echelon 4) is 1%, the accuracy required at Echelon 3, Echelon 2, and Echelon 1, respectively, is
(a) 0.1%, 0.01%, and 0.001%.
(b) 0.001%, 0.01%, and 0.1%.
(c) 0.1%, 0.001%, and 0.01%.
(d) 0.001%, 0.1%, and 0.01%.

6. The inaccuracy in a measurement may be due to
(a) the performance of equipment used to make the measurement.
(b) the test procedure.
(c) environmental effects.
(d) any or all of the above factors.

7. The uncertainty statement such as ± 0.1 units, 2 sigma confidence level means that
(a) it is 98% sure that the true value is not different from the measured value by more than ± 0.1 units.
(b) it is 95% sure that the true value is not different from the measured value by more than ± 0.1 units.
(c) it is 95% sure that the true value is different from the measured value by more than ± 0.1 units.
(d) it is 98% sure that the true value is different from the measured value by more than ± 0.1 units.

8. A company manufacturing current meters has set internal goals for achieving 4 to 1 test uncertainty ratios for all calibration tests. The tolerances of two meters (meter 1 and meter 2) respectively are ±0.1 mA and ±0.06 mA. The uncertainty of the calibration equipment is ±0.02 at a 95% confidence level. As per the goals set up by the company,
(a) meter 1 is acceptable and meter 2 is not acceptable.
(b) both meters are acceptable.
(c) none of the meters is acceptable.
(d) meter 2 is acceptable and meter 1 is not acceptable.

9. A Josephson junction is driven by a microwave source of 70 GHz. The junction is biased for quantum state (n) = 1. The voltage produced by the junction is about
(a) 1 mV
(b) 10 mV
(c) 145 μV
(d) 145 mV

10. Quantum Hall effect is used for realizing the primary
(a) frequency standard.
(b) resistance standard.
(c) current standard.
(d) time standard.

11. Single electron tunneling (SET) is used for realizing the primary
(a) frequency standard.
(b) resistance standard.
(c) current standard.
(d) voltage standard.

12. Single electron tunneling (SET) is used for realizing the primary
(a) capacitor standard.
(b) resistance standard.
(c) frequency standard.
(d) voltage standard.

13. It is expected that eventually all base units in metrology will be traceable back to time/frequency because
(a) no other standard can presently be measured so accurately.
(b) high level of accuracy and stability can be obtained relatively inexpensively.
(c) it can be communicated via electromagnetic waves.
(d) of all of the above factors.

Chapter 13

1. The purpose of ISO 9000 is to
(a) harmonize a large number of national/international standards.
(b) install and maintain a Quality System to strengthen competitiveness.
(c) establish Quality System standards in place of product standards.
(d) implement all of the above.

2. Which of the following ISO 9000 does not dictate to
(a) become globally integrated.
(b) set up product specifications.
(c) avoid multi third-party inspections/cost overruns.
(d) establish transparency of Quality Systems through evidence/documentation.

3. ISO 9001:1994, ISO 9002:1994 and ISO 9003:1994 are the core Quality Systems documents, which represent model for Quality Assurance, respectively, in
(a) design and development, production, installation and servicing, and final inspection and testing.
(b) production, installation and servicing, design and development, and final inspection and testing.
(c) production, installation and servicing, final inspection and testing, and design and development.
(d) final inspection and testing, production, installation and servicing, and design and development.

4. Environmental Management Standards are represented by
(a) ISO 9000 series.
(b) ISO 10000 series.

(c) ISO 14000 series.
(d) QS 9000 series.

5. Standards specifically developed for testing and calibration laboratories are represented by
(a) ISO 10011.
(b) ISO 17025.
(c) ISO 8402.
(d) ISO 10013.

6. Quality planning covers
(a) product planning.
(b) managerial and operation planning.
(c) preparation of quality plans and provisions for quality improvement.
(d) all of the above.

7. The goal of Quality Assurance of an industry is to develop its image and reputation with regard to
(a) product quality and reliability.
(b) price and delivery.
(c) customer care.
(d) all of the above.

8. The sequence of Deming cycle activities is:
(a) Do, Plan, Check, Act.
(b) Plan, Do, Check, Act.
(c) Act, Plan, Do, Check.
(d) Check, Plan, Do, Act.

9. Which of the following is not an activity associated with the customer/consumer in the Quality loop?
(a) Packaging and Storage.
(b) Installation and Operation.
(c) Technical Assistance and Maintenance.
(d) Sales and Distribution.

10. Which of the following is not an activity associated with the producer/supplier in the Quality loop?
(a) Product Design and Development.
(b) Process Planning.
(c) Disposal after use.
(d) Inspection and Testing.

11. Which of the following is not a component of Quality cost?
(a) Prevention cost
(b) Installation cost
(c) Appraisal cost
(d) Cost of failure

12. As the Quality becomes an integral element, which of the following is not correct?
   (a) Preventive and appraisal costs stabilize to the zero defect point.
   (b) Failure costs decline as preventive and appraisal costs level out.
   (c) Total cost is maximum at the zero defect point.
   (d) Total cost decreases to the zero defect point.

13. Which of the following is not a part of Quality audit?
   (a) Not including all activities within the company's normal operating procedure.
   (b) Examination of procedures.
   (c) Examination of records and results.
   (d) Discussions with staff.

14. Hierarchy of a Quality System from minimum Quality to zero defect Quality is
   (a) Quality Threshold, ISO Quality System, Total Quality Management, World Class Quality.
   (b) ISO Quality System, Quality Threshold, Total Quality Management, World Class Quality.
   (c) Total Quality Management, Quality Threshold, ISO Quality System, World Class Quality.
   (d) World Class Quality, Quality Threshold, ISO Quality System, Total Quality Management.

15. The key changes in ISO 9000:2000 vis-à-vis ISO 9000:1994 are with reference to
   (a) customer satisfaction.
   (b) continual improvement.
   (c) resource management.
   (d) all of the above.

Chapter 14

1. The sequence of seven layers of OSI model is:
   (a) Physical, Data Link, Network, Transport, Presentation, Session, Application.
   (b) Physical, Data Link, Network, Transport, Session, Presentation, Application.
   (c) Physical, Data Link, Network, Session, Transport, Presentation, Application.
   (d) Physical, Data Link, Network, Presentation, Transport, Session, Application.

2. RS and C in RS 232C respectively stand for:
   (a) Recommended Standard and Cable.
   (b) Run Serial and Cable.
   (c) Recommended Standard and third revision.
   (d) Run Serial and third revision.

3. Which of the following has differential mode of operation?
   (a) RS 422 and RS 485
   (b) RS 423 and RS 485
   (c) RS 232 and RS 422
   (d) RS 232 and RS 423

4. The allowed voltages in RS 232 are:
   (a) Logic 1 > +3 V and logic 0 < –3 V.
   (b) Logic 1 < –3 V and logic 0 > +3 V.
   (c) Logic 1 < –12 V and logic 0 > +12 V.
   (d) Logic 1 > +12 V and logic 0 < –12 V.

5. The cable length in RS 232
   (a) decreases as the baud rate decreases.
   (b) is independent of the baud rate.
   (c) is important at very low frequencies operation.
   (d) decreases as the baud rate increases.

6. Bluetooth technology
   (a) is a short-range wireless communication protocol.
   (b) provides wireless RS 232 interface.
   (c) operates in 2.4 GHz to 2.5 GHz frequency band.
   (d) includes all of the above.

7. Which of the following is not true for RS 422?
   (a) It provides full duplex communication.
   (b) Its communication range distance is higher than that of RS 232.
   (c) It is operated in single-mode communication mode.
   (d) Its communication speed is higher than that of RS 232.

8. Which of the following properties of RS 485 is not supported by RS 422?
   (a) Differential interface.
   (b) Use of the twisted pair cable.
   (c) Tristate capability.
   (d) Baud rate.

9. In an IEEE-488 bus, an instrument cannot be a
   (a) talker only.
   (b) listener only.
   (c) both talker and listener.
   (d) talker, listener, and controller simultaneously.

10. Which of the following is not true for the IEEE-488 bus?
   (a) It has 21 lines.
   (b) It provides a multi-drop interface.
   (c) It provides a parallel interface.
   (d) It provides digital communication.

11. IEEE-488 bus uses
   (a) TTL Logic.
   (b) CMOS Logic.
   (c) ECL Logic.
   (d) none of the above.

12. Which of the following is not true for the HS-488 bus?
   (a) Its baud rate is higher than that of IEEE-488.
   (b) It is slower than that of IEEE-488.
   (c) It is implemented in hardware.
   (d) It requires no changes in the software written for IEEE-488.

13. Which of the following is not true for the HART protocol?
   (a) It maintains compatibility with the 4–20 mA system.
   (b) It uses no digital communication.
   (c) It is a master/slave protocol.
   (d) It uses twisted pair cables.

14. A Bridge belongs to the following layers of the OSI model.
   (a) Physical and Data Link
   (b) Network and Data Link
   (c) Physical and Network
   (d) Physical, Data Link, and Network.

15. A Repeater encodes and decodes signals only at the following layer of the OSI model.
   (a) Network
   (b) Data Link
   (c) Physical
   (d) Application

16. Which of the following is not the property of Gigabit Ethernet?
   (a) Its data rate is 1000 Mbps.
   (b) It provides higher single-mode fiber cable length than that provided by Ethernet.
   (c) It provides smaller multi-mode fiber cable length than that provided by Ethernet.
   (d) It provides smaller STP/coaxial cable length than that provided by Ethernet.

17. A Router belongs to the following layers of the OSI model.
   (a) Physical, Data Link, and Transport.
   (b) Physical, Network, and Transport.
   (c) Physical, Data Link, and Network.
   (d) Data Link, Network, and Transport.

18. Which of the following features is not provided by a Fieldbus?
   (a) Bi-directional digital communication
   (b) Multi-variable access
   (c) Remote configuration
   (d) Increase in wiring

19. The use of a Fieldbus results in
   (a) higher reliability operation.
   (b) virtually unlimited flexibility.
   (c) reduction in equipment and installation costs.
   (d) all of the above.

20. Foundation Fieldbus uses the following layers of the OSI model.
   (a) Physical, Data Link, and Transport
   (b) Physical, Data Link, and Application
   (c) Data Link and Application
   (d) Physical, Presentation and Application

21. The communication rate provided by Foundation Fieldbus can be
   (a) 31.25 kbps.
   (b) 1 Mbps.
   (c) 2.5 Mbps.
   (d) all of the above.

Chapter 15

1. In a hazardous area, fire or explosion could occur when
   (a) combustible material is present.
   (b) combustible material is mixed with air.
   (c) a source of ignition ignites the mixture.
   (d) all of the above are present.

2. Which of the following does not produce ignition?
   (a) Discharge of capacitive circuits
   (b) Low temperature sources
   (c) Interrupting (opening) of inductive circuits
   (d) Opening or closing of resistive circuits with slow intermittent interruption

3. Which of the following approach(es) can be used to prevent ignition?
   (a) Intrinsically safe/Nonincendive equipment
   (b) Gas purge/Hermetic sealing
   (c) Flameproof/Explosion-proof housing
   (d) Any or all of the above

4. Textile mills fall in the following class of hazardous area classification.
   (a) Class I
   (b) Class II
   (c) Class III
   (d) None of the above

5. Petroleum refineries fall in the following class of hazardous area classification.
   (a) Class I
   (b) Class II
   (c) Class III
   (d) None of the above

6. Ignition energy required for wheat and acetylene respectively is
   (a) 240 mW/s, 0.017 mW/s
   (b) 0.017 mW/s, 0.017 mW/s
   (c) 240 mW/s, 240 mW/s
   (d) 0.017 mW/s, 240 mW/s

7. IP and NEMA ratings are used for
   (a) specifying the regulated power supply.
   (b) defining the protection level of the devices.
   (c) designing integrated circuits.
   (d) defining the protection level of the enclosures.

8. The first and the second characteristic numerals in IP rating, respectively, denote
   (a) protection against solid objects and protection against gases.
   (b) protection against solid objects and protection against liquids.
   (c) protection against liquids and protection against solid objects.
   (d) protection against gases and protection against solid objects.

9. Which of the following is not true for the NEMA definitions?
   (a) Definitions cover indoor enclosures
   (b) Definitions cover outdoor enclosures
   (c) Definitions do not cover enclosures for hazardous areas
   (d) Definitions cover protection from dust

10. Pressurized enclosures provide
   (a) thermal/convection cooling.
   (b) near-ambient cooling.
   (c) below-ambient cooling.
   (d) any of the above.

11. The components of an intrinsically safe circuit include
   (a) field device.
   (b) intrinsically safe barrier.
   (c) field wiring.
   (d) all of the above.

12. The barrier contains
   (a) a resistor
   (b) zener diodes
   (c) a fuse
   (d) all of the above

13. Which of the following is not a selection parameter for a barrier?
   (a) Ignition temperature
   (b) Rated voltage
   (c) Polarity
   (d) Internal resistance

14. Which of the following should be considered while installing an intrinsically safe system?
   (a) Proper wiring arrangement
   (b) Proper grounding arrangement
   (c) Proper sealing arrangement
   (d) All of the above

15. Which of the following is not the requirement of a Transformer Isolated Barrier?
(a) High integrity intrinsically safe ground
(b) Zener diodes
(c) Current limiting circuitry
(d) Unregulated power supply

Answers to Multiple Choice Questions

Chapter 1

1. (b), **2.** (c), **3.** (d), **4.** (c), **5.** (d), **6.** (b), **7.** (c), **8.** (c), **9.** (c), **10.** (a), **11.** (d), **12.** (a), **13.** (b), **14.** (d), **15.** (c), **16.** (a), **17.** (c), **18.** (d), **19.** (c), **20.** (a)

Chapter 2

1. (c), **2.** (a), **3.** (b), **4.** (a), **5.** (d), **6.** (a), **7.** (d), **8.** (b), **9.** (b), **10.** (d), **11.** (c), **12.** (d), **13.** (b), **14.** (b), **15.** (c)

Chapter 3

1. (c), **2.** (b), **3.** (b), **4.** (c), **5.** (c), **6.** (d), **7.** (a), **8.** (a), **9.** (c), **10.** (d), **11.** (c), **12.** (a), **13.** (b), **14.** (c), **15.** (b), **16.** (d), **17.** (b), **18.** (a)

Chapter 4

1. (d), **2.** (a), **3.** (c), **4.** (d), **5.** (b), **6.** (a), **7.** (c), **8.** (c), **9.** (d), **10.** (c), **11.** (b), **12.** (b), **13.** (a), **14.** (c), **15.** (c)

Chapter 5

1. (d), **2.** (d), **3.** (d), **4.** (d), **5.** (a), **6.** (b), **7.** (c), **8.** (a), **9.** (a), **10.** (a), **11.** (c), **12.** (d), **13.** (a), **14.** (c), **15.** (d), **16.** (a), **17.** (b), **18.** (d)

Chapter 6

1. (d), **2.** (b), **3.** (a), **4.** (d), **5.** (c), **6.** (b), **7.** (a), **8.** (c), **9.** (c), **10.** (d)

Chapter 7

1. (c), **2.** (d), **3.** (c), **4.** (c), **5.** (b), **6.** (b), **7.** (b), **8.** (a), **9.** (c), **10.** (d), **11.** (a), **12.** (c), **13.** (d), **14.** (b), **15.** (c), **16.** (b), **17.** (c), **18.** (c), **19.** (c), **20.** (a), **21.** (d), **22.** (b), **23.** (a), **24.** (b)

Chapter 8

1. (a), **2.** (b), **3.** (b), **4.** (c), **5.** (c), **6.** (c), **7.** (a), **8.** (b), **9.** (c), **10.** (a), **11.** (c), **12.** (a), **13.** (c), **14.** (d), **15.** (b), **16.** (b), **17.** (c), **18.** (c), **19.** (b), **20.** (a), **21.** (b), **22.** (d)

Chapter 9

1. (b), **2.** (a), **3.** (d), **4.** (b), **5.** (c), **6.** (d), **7.** (a), **8.** (c), **9.** (b), **10.** (d), **11.** (b), **12.** (c), **13.** (d), **14.** (a), **15.** (c), **16.** (a), **17.** (b), **18.** (c), **19.** (a), **20.** (b)

Chapter 10

1. (c), **2.** (c), **3.** (d), **4.** (b), **5.** (c), **6.** (c), **7.** (d), **8.** (d), **9.** (a), **10.** (d), **11.** (d)

Chapter 11

1. (b), **2.** (d), **3.** (d), **4.** (b), **5.** (a), **6.** (d), **7.** (b), **8.** (d), **9.** (b), **10.** (c), **11.** (d), **12.** (d), **13.** (a), **14.** (d), **15.** (a)

Chapter 12

1. (d), **2.** (c), **3.** (a), **4.** (b), **5.** (a), **6.** (d), **7.** (b), **8.** (a), **9.** (c), **10.** (b), **11.** (c), **12.** (a), **13.** (d)

Chapter 13

1. (d), **2.** (b), **3.** (a), **4.** (c), **5.** (b), **6.** (d), **7.** (d), **8.** (b), **9.** (a), **10.** (c), **11.** (b), **12.** (c), **13.** (a), **14.** (a), **15.** (d)

Chapter 14

1. (b), **2.** (c), **3.** (a), **4.** (b), **5.** (d), **6.** (d), **7.** (c), **8.** (c), **9.** (d), **10.** (a), **11.** (a), **12.** (b), **13.** (b), **14.** (a), **15.** (c), **16.** (b), **17.** (c), **18.** (d), **19.** (d), **20.** (b), **21.** (d)

Chapter 15

1. (d), **2.** (b), **3.** (d), **4.** (c), **5.** (a), **6.** (a), **7.** (d), **8.** (b), **9.** (c), **10.** (d), **11.** (d), **12.** (d), **13.** (a), **14.** (d), **15.** (a).

Bibliography

Bentley, John P., *Principles of Measurement Systems,* Addison Wesley Longman, 2000.

Bolton, W., *Industrial Control & Instrumentation,* Orient Longman, New Delhi, 1993.

Cain, W.D., *Engineering Product Design*, Business Books Limited, 1969.

Car Joseph, J., *Elements of Electronic Instrumentation and Measurement,* 2nd ed., Prentice Hall Inc., Englewood Cliffs, New Jersey, 1986.

Coombs, Clide F. Jr., *Electronic Instrument Handbook,* McGraw-Hill, New York, 1994.

Daugherty, K.M., *Analog-to-Digital Conversion,* McGraw-Hill, New York, 1995.

David, W. Allan, Neil Ashby, and Clifford C. Hodge, *The Science of Timekeeping,* Application Note 1289, Hewlett Packard, 1997.

Finkelstein, L. and K.T.V., Grattan, *Concise Encyclopedia of Measurement & Instrumentation*, Pergamon Press, New York, 1994.

Fowler, Kim R., *Electronic Instrument Design,* Oxford University Press, 1996.

Harvey, Glen F., *Standards and Practices for Instrumentation,* Instrument Society of America, 1974.

Heiserman, David L., *Surface Mount Devices: Troubleshooting and Repair,* Prentice Hall Inc., Englewood Cliffs, New Jersey, 1990.

IEEE Transactions on Instrumentation and Measurement, Vol. **18**, No. 2, April 1999. (a number of articles).

Kaplan Gadi, "On Good Design", *IEEE Spectrum*, May 1987, pp. 27-28.

Larry Jones and Chin A. Foster, *Electronic Instrument and Measurements,* John Wiley & Sons, New York 1983.

Malvino A.P., *Electronic Instrumentation Fundamentals,* McGraw-Hill, New York, 1967.

Mark W. Keller, "Standards of Current and Capacitance Based on Single-Electron Tunneling Devices", *Proceedings of Fermi School CXLVI: Recent Advances in Metrology and Fundamental Constants*, October 2000.

Morris, Allen S., *Principles of Measurement and Instrumentation,* 2nd ed., Prentice-Hall of India, New Delhi 1990.

Ostrifsky Benjamin, *Design, Planning and Development Methodology*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1977.

Pohlmann, Ken, C., editor-in-chief, *Advanced Digital Audio,*1st ed., Sams, 1991.

Prensky, Sol D. and L. Castellucis Richard, *Electronic Instrumentation,* 3rd ed., Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1982.

Ray, Martin S., *Elements of Engineering Design*, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1985.

Rechtin Eberhardt, "The Art of System Architecting", *IEEE Spectrum*, Oct. 1992, pp. 66–69.

Reddding, R.J., *Intrinsic Safety; The Safe Use of Electronics in Hazardous Locations,* McGraw-Hill, New York, 1971.

Skytte Kurt, "Engineering a Small System", *IEEE Spectrum*, Mar. 1994, pp. 63–65.

Traister Johan E., *Design Guidelines for Surface Mount Technology,* Academic Press, 1990.

Watkinson John, *The Art of Digital Audio,* 3rd ed., Focal Press, Butterworth Group, 1989.

Woodson, Wesley E. and W. Conover Donald, *Human Engineering Guide for Equipment Designers,* 2nd ed., University of California Press, 1973.

Answers to Numerical Questions

Chapter 1

1.4 100 A

1.7 Ammeter resistance = 10.02 Ω
Voltmeter resistance = 998 kΩ

1.8 With sensitivity = 100 kΩ/*V*, voltmeter reading = 3.33 V, % error = 33.4%
With sensitivity = 500 kΩ/V, voltmeter reading = 4.55 V, % error = 9%

1.9 (a)Ratio = $\dfrac{R_Z + R_m}{R_x + R_Z + R_m}$
(b) For 20% fsd, R_x = 12 kΩ
For 50% fsd, R_x = 3 kΩ
For 100% fsd, R_x = 0 Ω

1.10 (a) R_2 = 125 Ω, R_1 = 4944.4 Ω
(b) Change in R_2 = 36.29 Ω
(c) R_h = 5006.17 Ω

1.11 R_h = 1.01 Ω, R_1 = 29 Ω

1.12 8.26 V

1.13 (c)R_1 = 3420 Ω, R_2 = 14,400 Ω, R_3 = 18,000 Ω
(d) 360 Ω/V

1.15 Reading of M_1 = 0.8 A
Reading of M_2 = 552 V

1.16 R = 300 Ω, V_{DC} = 1.33 V

1.17 (a)p-to-p reading = 10 V, % error = 0%
(b)rms indication = 6.5 V, % error = 12.7%
(c) rms indication = 7.07 V, % error = 22.5%
(d)rms indication = 5.77 V, % error = 0%

1.18 C = 10 μF, max. peak value which can be measured = 10 V

1.19 Meter will read high.

1.20 See Figure 1.37

Chapter 2

2.3 (b) 10101101

2.5 0.5 V

2.7 (b) 200 pA

2.8 156,863 conversions per second

Chapter 4

4.5 36 pF

4.7 445 MHz, 1 GHz
445 MHz will be visible in the measured waveform whereas 1 GHz will not be visible in the measured waveform.

4.8 21.21 MHz

4.10 3000:1

Chapter 5

5.1 20 μΩ

5.5 0.00032

5.11 The bridge is balanced

5.14 $\Delta 6$

5.16 $\Delta R_4 = 2150\ \Omega$, $C_4 = 95.7$ nF
$\Delta R_4 = \pm 10\ \Omega$, $\Delta C_4 = \pm 0.3$ nF

5.17 $\Delta R_x = 10\ \text{k}\Omega$, $X_x = 22.8 \times 10^{-6}\ \Omega$
$\Delta L_x = 3.63 \times 10^{-13}$ H

Chapter 7

7.3 2 kΩ, 1.6 kHz

7.4 $R = 6.5\ \text{k}\Omega$, $R_1 = 2\ \text{k}\Omega$, $R_2 = 58\ \text{k}\Omega$

7.6 $C = 50.66$ pF. Let $R_1 = R_2 = 1\ \text{k}\Omega$

7.7 $C_1 = C_2 = 101.32$ pF. Let $R_1 = R_2 = 1\ \text{k}\Omega$

7.8 $f_s = 7.955$ MHz, $f_a = 8.013$ MHz,
$Q = 36{,}882$

7.12 $R_1 = R_2 = 10\ \text{k}\Omega$

7.13 Characteristics will be noninverting

7.14 $C = 0.01\ \mu$F. Since $T_1 = T_2$, waveform is a square wave.

7.15 $R_2 = 20\ \text{k}\Omega$, $R = 50\ \text{k}\Omega$

7.16 $R = 5.78\ \text{k}\Omega$

7.22 Refer to Figure 7.32. $m = 2$, $n = 7$ and $\Delta f = 1$ MHz

7.23 Refer to Figure 7.36. $N_3 = 4$, $N_2 = 2$, $N_1 = 3$

Chapter 8

8.11 (b) 80 dB

Chapter 9

9.13 (b) Less than 100 kHz

9.14 Total time base error dominates over ±1 count error

9.21 ±0.01 Hz

9.22 ±50.30 ps

9.25 ±17.4 Hz

Chapter 11

11.4 Range of measurement: 0–100 mV DC
Accuracy: 0.01%
Stability: 1 digit
Operating temp. range: 20–40°C
Operating RH: 35–65%

11.24

1. Unstable and fluctuating ac input voltage
2. Environmental temperature greater than 45°C
3. Ambient RH greater than 65%
4. Improper grounding and shielding

Chapter 12

12.6 (a) 10 ppm, NPL, New Delhi
(b) 0.25%, ETDC, Jaipur
(c) 2.5 ppm, NPL, New Delhi
(d) 0.25%, ETDC, Chennai

Appendix A

A.1 $R_A = 200\ \text{k}\Omega$, $R_B = 600\ \text{k}\Omega$,
$R_C = 1.2\ \text{M}\Omega$, $R_D = 6\ \text{M}\Omega$

A.2 Two attenuators in cascade, each with $R_1 = 409\ \Omega$, $R_2 = 101\ \Omega$

A.3 $R^0 = 150\ \Omega$, $V_{out} = 253.56$ mV

A.4 $R_1 = 9.48\ \text{k}\Omega$, $R_2 = 639\ \Omega$,

A.5 59.1 Ω, 64.3 Ω

Index

AFG. *See* Signal generator
Analog meters, 3
  ac analog meters, 28
  calibration, 27
  d'Arsonval movement, 6
  dc ammeter, 10
  dc analog meters, 10
  dc voltmeter, 16
  dynamic behaviour, 8
  electrodynamic meter movement, 38
  electronic analog
    ac ammeter, 47
    ac voltmeter, 45
    dc ammeter, 45
    dc voltmeter, 43
    meters, 43
    multimeter, 48
    ohmmeter, 47
  four-wire sensing, 48
  full wave rectifier based meter, 32
  half wave rectifier based meter, 28
  moving iron (vane) meter movement, 41
  multimeter, 25
  nonsinusoidal waveform input, 33
  ohmmeter, 20
  peak reading voltmeter, 34
  peak-to-peak reading voltmeter, 37
  permanent magnet moving coil, 3
  two-wire sensing, 47
ARB. *See* Signal generator
Attenuator, 599
  basics, 599
  bridged T, 603
  characteristic resistance, 600
  padding, 605
  symmetrical π, 602
  symmetrical T, 600
  variable, 603
AWG. *See* Signal generator

Bridge instruments, 179
  ac bridges, 200
  Anderson loop, 212
  dc bridges, 179
  Hay's bridge, 206
  Kelvin bridge, 195
  Kelvin double bridge, 196
  LCR bridge, 218
  Maxwell bridge, 205
  Schering bridge, 208
  Varley loop test, 185
  Wagner ground connection, 211
  Wein bridge, 210
  Wheatstone bridge, 180

Data logger. *See* Recorders
DDS. *See* Signal generator
DG. *See* Signal generator
Digital meters, 54
  aliasing, 75
  analog-to-digital converter, 66
  digital meter displays, 92
  digital-to-analog converter, 55
  digital ramp A/D converter, 70
  DMM calibration, 105
  dual slope A/D converter, 71
  inverted *R*-2*R* ladder D/A converter, 60
  oversampling, 82

- parallel or flash A/D converter, 79
- performance characteristics of
  - D/A converters, 61
  - A/D converters, 87
- *R*-2*R* ladder D/A converter, 58
- sample-hold circuits, 77
- selecting a DMM, 112
- sigma delta A/D converter, 80
- single slope (Ramp) A/D converter, 69
- specifications of, 92
- successive approximation A/D converter, 73
- switched-capacitor D/A converter, 60
- undersampling, 77
- weighted-resistor D/A converter, 57

DPO. *See* Digital phosphor oscilloscope
DSO. *See* Digital storage oscilloscope

Electronic counters, 338
- arming, 357
- conventional electronic counter, 339
- design considerations, 344
- frequency measurement, 339
- frequency measurement error, 352
- frequency ratio measurement, 344
- interpolation, 356
- period measurement, 342
- period measurement error, 352
- reciprocal electronic counters, 354
- specifications of, 359
- time interval measurement, 343
- time interval measurement error, 354
- totalizer, 344
- types of measurement error, 350

Elements of design, 389
- aesthetics, 415
- assembly and inspection, 408
- availability, 398
- circuit design, 394
- circuit layout, 404
- cooling, 412
- deployment, 393
- documentation, 422
- enclosures, 411
- ergonomics, 417
- fault tolerance, 399
- high speed design, 401
- input interface, 404
- integrated testing, 420
- low power design, 403
- manufacturing process design, 392
- MDT, 399
- MTBF, 399
- MTTF, 399
- MTTR, 399
- MUT, 399
- noise and error budgets, 403
- output interface, 404
- power distribution, 409
- product design, 391
- product life cycle, 390
- regulations and standards, 420
- reliability, 397
- retirement/disposal, 394
- selection of technology, 397
- serviceability, 420
- system engineering, 389
- testing and calibration, 408
- upgrades, 393
- wiring and cabling, 410

ESD. *See* Grounding and shielding

Grounding and shielding, 371
- basic grounding practices, 376
- considerations, 376
- electrostatic discharge, 384
- grounding, 371
- guidelines, 382
- shielding, 382
- shock hazard, 375

Industrial communication techniques, 497
- backbones, 536
- bluetooth technology, 510
- bridge, 531
- CIM pyramid, 538
- CSMA/CD, 533
- Ethernet, 533
- Fieldbus, 538
- Foundation Fieldbus, 548
- GPIB, 515
- HART, 527
- HS 523
- IEEE 488, 515
- IEEE 488.2, 521
- network topologies, 502
- OSI network model, 500
- physical layer options, 502
- repeater, 532
- router, 530
- RS 232, 503
- RS 422, 511
- RS 423, 513
- RS 485, 513
- token buses and rings, 531
- wireless RS 232, 510

Instrumentation in hazardous areas, 564
- barrier, 588
- CENELEC, 565

- enclosure classification designations, 569
- equipment design and construction, 575
- explosion-proof, 569
- flameproof, 569
- gas purge, 569
- hazardous area classifications, 565
- hazardous areas, 564
- hermetic sealing, 569
- IEC, 565
- ignition triangle, 564
- ingress protection (IP), 570
- intrinsically safe, 578
- NEMA, 571
- non-incendive, 569
- protective concepts, 567
- transformer isolation, 593

Instrumentation symbols, 619

Metrology, 426
  - calibration, 431
  - calibration programme, 432
  - capacitance standard, 452
  - compatibility, 431
  - current standard, 450
  - electrical standards, 443
  - Global Positioning System (GPS), 454
  - inductance standard, 453
  - Josephson array voltage standard, 443
  - measurement glossary, 441
  - measurement reliability, 439
  - measurement uncertainty, 435
  - measuring equipment cycle, 434
  - quantum Hall effect, 447
  - resistance standard, 446
  - SI units, 428
  - Single Electron Tunneling (SET), 450
  - test uncertainty ratio, 437
  - testing, 430
  - time and frequency standard, 453
  - traceability, 432
  - traceability echelons, 434
  - voltage standard, 443

Oscilloscope, 119
  - analog, 122
  - controls, 138
  - CRT, 125, 133
  - digital phosphor, 137
  - digital storage, 132, 137
  - DPO, 137
  - electromagnetic deflection, 133
  - equivalent-time sampling, 137
  - measurements, 151
  - phosphor material, 128
  - random scan, 131
  - raster scan, 131
  - real-time sampling, 136
  - sampling methods, 135
  - sin *x*/*x* interpolation, 136
  - trigger, 122, 144
  - Z modulation, 129

P & I diagrams, 619

PLL. *See* Signal generator

Probes, 159
  - 50 ohm divider, 160
  - active, 163
  - active (FET) voltage, 163
  - ac/dc current, 165
  - current, 164
  - differential, 163
  - high Voltage, 162
  - loading and measurement effects, 167
  - logic, 166
  - optical, 166
  - passive voltage, 160
  - specifications, 173
  - types, 160

Real-time sampling, 136

Recorders, 226
  - accessories, 235
  - charts, 235
  - data logger, 237
  - digital, 237
  - distinction between a data logger, 227
  - dynamic behaviour, 230
  - hybrid recorders, 227
  - null balancing, 232
  - paperless, 238
  - PMMC galvanometer type, 228
  - potentiometric, 232
  - selection, 239
  - servo, 232
  - strip chart, 228
  - ultrasonic pen position sensing, 233
  - videographic, 238
  - writing mechanisms, 229
  - *X-Y*, 236

Signal analyzer, 295
  - distortion, 296
  - distortion analyzer, 298
  - DSO analyzer, 316
  - FFT analyzer, 311

  - IMD, 297
  - interfacing target system, 334
  - intermodulation analyzer, 300
  - logic analyzer, 318
  - logic state analyzer, 328
  - logic timing analyzer, 320
  - SINAD, 297
  - spectrum analyzer, 303
  - THD, 296
  - vector analyzer, 315
  - wave analyzer, 300
- Signal generator, 241
  - arbitrary function generator, 288
  - arbitrary waveform generator, 283
  - astable multivibrator, 258
  - audio freq. signal generator, 244
  - bistable multivibrator, 255
  - Colpitts oscillator, 250
  - crystal oscillator, 250
  - data generator, 289
  - digital signal generators, 282
  - direct analog synthesis, 270
  - direct digital synthesis, 277
  - fractional–N indirect synthesizer, 274
  - frequency synthesis techniques, 269
  - function generator, 264
  - Hartley oscillator, 249
  - indirect synthesis, 272
  - key characteristics, 290
  - monostable multivibrator, 262
  - multiple loop indirect synthesizer, 273
  - nonsinusoidal generator, 254
  - PLL, 272
  - programmable decade indirect synthesizer, 273
  - pulse parameter definitions, 265
  - RC phase shift oscillator, 246
  - RF signal generator, 247
  - Schmitt trigger, 256
  - sine wave generator, 242
  - standard frequency synthesis, 274
  - VCO, 272
  - Wein bridge oscillator, 245
- Standards in quality management, 457
  - Deming cycle, 468
  - E-base, 486
  - elements of ISO 9001, 474
  - ISO 10005, 461
  - ISO 10007, 461
  - ISO 10011, 463
  - ISO 10012, 463
  - ISO 10013, 463
  - ISO 14000, 484
  - ISO 14001, 484
  - ISO 17025, 485
  - ISO 8402, 463
  - ISO 9000, 458
  - 2000 update, 487
  - ISO 9001, 460
  - ISO 9002, 460
  - ISO 9003, 460
  - ISO 9004, 462
  - Japanese Industrial Standards (JIS), 485
  - Q-base, 486
  - QS 9000, 485
  - Quality
    - assurance, 467
    - audits, 472
    - control, 466
    - cost, 469
    - definition, 465
    - improvement, 471
    - loop, 469
    - management, 466
    - policy, 466
    - system, 466
    - tools, 473
  - seven wastes, 471

- Total Quality Management (TQM), 481

Electronic Instruments and Instrumentation Technology

M.M.S. Anand

The standard laboratory tools in the modern scientific world include a wide variety of electronic instruments used in measurement and control systems. This book provides a firm foundation in principles, operation, design, and applications of electronic instruments. Commencing with electromechanical instruments, the specialized instruments such as signal analyzers, counters, signal generators, and digital storage oscilloscope are treated in detail.

Good design practices such as grounding and shielding are emphasized. The standards in quality management, basics of testing, compatibility, calibration, traceability, metrology and various ISO 9000 quality assurance guidelines are explained as well.

The evolution of communication technology in instrumentation is an important subject. A single chapter is devoted to the study of communication methods used in instrumentation technology.

There are some areas where instrumentation needs special type of specifications—one such area is hazardous area. The technology and standards used in hazardous areas are also discussed.

An instrumentation engineer is expected to draw and understand the instrumentation drawings. An Appendix explains the symbols and standards used in P&I diagrams with several examples.

Besides worked-out examples included throughout, end-of-chapter questions and multiple choice questions are also given to judge the student's understanding of the subject.

Practical and state-of-the-art in approach, this textbook will be useful for students of electrical, electronics, and instrumentation engineering.

ABOUT THE AUTHOR

M.M.S. ANAND, Ph.D., is Professor of Electronics and Instrumentation at Birla Institute of Technology & Science, Pilani. Presently, he also holds the position of Registrar of the Institute.

Prof. Anand has more than 13 years of teaching experience in electronics and instrumentation and 16 years of industrial experience. In the industry, he mostly worked on data acquisition and control systems. His current areas of interest include design and simulation of analog and digital electronic circuits.

ISBN:978-81-203-2454-1

9 788120 324541

www.phindia.com